



KADIR HAS UNIVERSITY  
SCHOOL OF GRADUATE STUDIES  
DEPARTMENT OF COMMUNICATION STUDIES

**COUNTERSPEECH AS A METHOD IN COMBATING  
ONLINE HATE SPEECH IN TURKEY: AN ONLINE  
SURVEY EXPERIMENT**

GÜLTEN OKÇUOĞLU

MASTER'S DEGREE THESIS

ISTANBUL, DECEMBER, 2022

Gülten Okçuođlu

Master's Degree Thesis

2023

**COUNTERSPEECH AS A METHOD IN COMBATING  
ONLINE HATE SPEECH IN TURKEY: AN ONLINE  
SURVEY EXPERIMENT**

GÜLTEN OKÇUOĞLU  
ADVISOR: ASST. PROF. ÖZEN BAŞ

MASTER'S DEGREE THESIS

A thesis submitted to the school of graduate studies with the aim to meet the partial requirements required to receive a master's degree in the department of communication studies

Istanbul, December, 2022

## APPROVAL

This thesis titled **COUNTERSPEECH AS A METHOD IN COMBATING ONLINE HATE SPEECH IN TURKEY: AN ONLINE SURVEY EXPERIMENT** submitted by **GÜLTEN OKÇUOĞLU**, in partial fulfillment of the requirements for the Master's Degree Thesis in Communication Studies is approved by

Asst. Prof., Özen Baş (Advisor) .....  
Kadir Has University

Asst. Prof., İrem İnceoğlu .....  
Kadir Has University

Asst. Prof., Suncem Koçer .....  
Koç University

I confirm that the signatures above belong to the aforementioned faculty members.

\_\_\_\_\_  
Prof. Dr., Mehmet Timur Aydemir  
Director of the School of Graduate Studies  
Date of Approval: 15/12/2022

## DECLARATION ON RESEARCH ETHICS AND PUBLISHING METHODS

I, GÜLTEN OKÇUOĞLU; hereby declare

- that this master's degree thesis that I have submitted is entirely my own work and I have cited and referenced all material and results that are not my own in accordance with the rules;
- that this master's degree thesis does not contain any material from any research submitted or accepted to obtain a degree or diploma at another educational institution;
- and that I commit and undertake to follow the "Kadir Has University Academic Codes of Conduct" prepared in accordance with the "Higher Education Council Codes of Conduct".

In addition, I acknowledge that any claim of irregularity that may arise in relation to this work will result in a disciplinary action in accordance with the university legislation.

Gülten Okçuoğlu

---

15.12.2022



*To My Dearest Family...*

## ACKNOWLEDGEMENT

Without the following people, I would not have been able to finish this thesis and would not have completed my master's degree.

Firstly, I would like to express my sincere gratitude for my supervisor Asst. Prof. Özen Baş. She has been an outstanding supervisor and I cannot thank Dr. Baş enough for her guidance, patience, sincerity and dedication throughout this journey. I appreciate so much her understanding when I was stressed and panicked. She has been not only a supervisor to me but also a mentor who has always supported and encouraged me to find new academic opportunities. I'm glad that we crossed paths with her.

After a ten-year break, I was filled with fear and anxiety when I decided to return to the academia. I would like to thank Asst. Prof. Suncem Koçer, who helped me in transforming this fear and anxiety into a great love. Working with Dr. Koçer in her project at the beginning of my master's program allowed me to understand scientific research in a more nuanced way. I would like to thank Asst. Prof. Dr. İrem İnceoğlu and also Dr. Koçer for being in my committee and for their valuable time. I look forward to their feedback.

Finally, I would like to express my sincere love for my dear family, particularly for my sister Dilan, who has been more than a sister to me. Although being physically apart, I have always felt her support. I thank Irmak Aytemür Bulut and Ozan Mete, my two close friends, who had to bear some of the anxiety or frustration one can feel during this journey. I am grateful for the patience, motivation and support my family and friends have been feeding me with. I feel blessed to have you all in my life.

# COUNTERSPEECH AS A METHOD IN COMBATING ONLINE HATE SPEECH IN TURKEY: AN ONLINE SURVEY EXPERIMENT

## ABSTRACT

With rapidly developing and changing technologies, online platforms transformed into a place where internet users can state their thoughts and opinions. Even though this feature represents a unique communication opportunity, it has also brought many strains alongside. One of these challenges is online hate speech, which can generate violence offline. Therefore, civil society organizations (CSOs), governments, institutions, social networks, and individuals create new strategies and methods to tackle it and reduce its effects. One of the potential solutions for this problem is counterspeech. As a method, the popularity of counterspeech has been enhancing daily. Yet, there is a dearth of research analyzing its effectiveness, particularly based on refugees. This thesis fills this gap by testing the effectiveness of counterspeech on Twitter in preventing people from posting hateful comments about refugees. Based on a survey experiment (N=181) conducted online in Turkey, with two conditions: (1) exposure to empathy-based counterspeech comments versus (counterspeech condition) (2) who were exposed to hateful comments (hate speech condition). Results suggest that there is a statistically significant relationship between these two conditions. In other words, the level of generating hostile messages in Twitter were higher among those who viewed hateful comments than those in the counterspeech condition. This result shows not only that counterspeech has an essential role in combating online hate speech but also will play a pioneering role in designing various hate speech combating methods developed by CSOs in Turkey in recent years.

**Keywords:** Online hate speech, hate speech, counterspeech, refugees, Twitter, online survey experiment, social media platforms



# TÜRKİYE'DE ÇEVİRİM İÇİ NEFRET SÖYLEMİYLE MÜCADELEDE BİR YÖNTEM OLARAK KARŞI SÖYLEM: ÇEVİRİMİÇİ ANKET DENEYİ

## ÖZET

Hızla gelişen ve değişen teknolojiyle beraber çevrim içi platformlar internet kullanıcılarının düşünce ve fikirlerini belirtebilecekleri bir alana dönüştü. Her ne kadar bu özellik eşsiz bir iletişim fırsatı sunuyor olsa da beraberinde birçok zorluğu da getirmektedir. Bu zorluklardan birisi de olumsuz etkileriyle çevrim dışı ortamda da karşılaştığımız çevrim içi nefret söylemidir. Bu sorunla mücadele etmek ve etkilerini azaltmak için sivil toplum kuruluşları (STK), hükümetler, sosyal ağ sağlayıcıları, kurumlar ve bireysel aktivistler yeni stratejiler ve yöntemler geliştirmektedir. Bu yöntemlerden birisi de karşı konuşmadır. Her ne kadar karşı konuşmanın popülaritesi her geçen gün artıyor olsa da özellikle mülteciler özelinde bu stratejinin etkisini ölçen çalışma sayısı oldukça kısıtlıdır. Bu çalışma ile Twitter'da mültecilere yönelik üretilen nefret söylemleriyle mücadelede karşı konuşma yönteminin etki bir yöntem olarak kullanılıp kullanılmayacağı test edilecektir. Çevrim içi anket deneyi ( $N=181$ ) yöntemi ile (1) mültecilerle ilgili empati ve yeniden insanlaştırma temelli karşı konuşmaya maruz kalanlarla (karşı konuşma koşulu) (2) mültecilerle ilgili nefret dolu yorumlara maruz kalanların yorum yapma davranışları analiz edilerek, karşı konuşmanın etkisi ölçülecektir. Sonuçlar, katılımcıların mültecilere yönelik var olan fikirleri kontrol altına alındığında, sadece karşı konuşma içeren yorumlara maruz kalanların, nefret söylemi içeren yorumlara maruz kalanlara oranla daha az nefret söylemi içeren yorum ürettiği gözlemlenmiştir. Bu sonuç, karşı konuşmanın sadece çevrim içi nefret söylemiyle mücadelede önemli bir rolü olduğunu değil aynı zamanda son yıllarda Türkiye'de faaliyet yürüten STK'lar tarafından sıklıkla gündemleştirilen çevrim içi nefret söylemiyle mücadele tasarımlarında karşı konuşmanın öncü bir role sahip olacağını göstermektedir.

**Anahtar Sözcükler:** Çevrim içi nefret söylemi, nefret söylemi, karşı konuşma, mülteciler, Twitter, çevrim içi anket deneyi, sosyal medya platformları

## TABLE OF CONTENTS

|  |                              |
|--|------------------------------|
| <b>ACKNOWLEDGEMENT</b> .....   | <b>v</b>                     |
| <b>ABSTRACT</b> .....  | <b>vi</b>                    |
| <b>ÖZET</b> .....  | Error! Bookmark not defined. |
| <b>LIST OF FIGURES</b> .....   | <b>x</b>                     |
| <b>LIST OF TABLES</b> .....  | <b>xi</b>                    |
| <b>LIST OF SYMBOLS</b> .....   | Error! Bookmark not defined. |
| <b>LIST OF ACRONYMS AND ABBREVIATIONS</b> .....                                    | <b>xi</b>                    |
| <b>1. INTRODUCTION</b> .....   | <b>1</b>                     |
| <b>2. LITERATURE</b> .....   | <b>5</b>                     |
| <b>2.1 Hate Speech</b> .....   | <b>5</b>                     |
| <b>2.1 Hate Speech</b> .....   | <b>9</b>                     |
| <b>2.3 Counterspeech</b> .....   | <b>12</b>                    |
| <b>2.3.1 Factors that affect the effectiveness of counterspeech</b> .....          | <b>14</b>                    |
| <b>3. METHODOLOGY</b> .....  | <b>18</b>                    |
| <b>3.1 Survey Experiment</b> .....   | <b>18</b>                    |
| <b>3.2 Design of the Experiment</b> .....  | <b>18</b>                    |
| <b>3.3 Experimental Stimuli</b> .....  | <b>23</b>                    |
| <b>3.4 Sample</b> .....  | <b>24</b>                    |
| <b>3.5 Variables</b> .....   | <b>26</b>                    |
| <b>3.5.1. Preexisting attitudes toward refugees</b> .....                          | <b>26</b>                    |
| <b>3.5.2. Dependent variable</b> .....   | <b>27</b>                    |
| <b>3.5.3. Independent variable</b> .....   | <b>29</b>                    |
| <b>4. RESULT</b> .....   | <b>30</b>                    |
| <b>4.1 Characteristics of the sample</b> .....                                     | <b>30</b>                    |
| <b>4.2 Effect of Counterspeech Condition</b> .....                                 | <b>30</b>                    |
| <b>4.3 The Effect of Political Party Voting on Writing Hate Speech</b> .....       | <b>31</b>                    |
| <b>4.4 The Effect of Ethnic Identity on Writing Hate Speech</b> .....              | <b>31</b>                    |
| <b>4.5. Analyzing Open-Ended Comments</b> .....                                    | <b>32</b>                    |
| <b>4.6 Analyzing Preexisting Perceptions of Participants Toward Refugees</b> ..... | <b>33</b>                    |

|   |           |
|---|-----------|
| <b>5. DISCUSSION .....</b>  | <b>34</b> |
| <b>BIBLIOGRAPHY .....</b>   | <b>38</b> |
| <b>APPENDIX A .....</b>   | <b>48</b> |
| <b>A.1 News and Comments Showed During the Experimental Stimuli .....</b> | <b>48</b> |
| <b>A.1.1 Hate Speech Condition .....</b>                                  | <b>48</b> |
| <b>A.1.2 Counterspeech Condition.....</b>                                 | <b>48</b> |
| <b>A.2 Questions to Measure Participants' Refugee Attitude.....</b>       | <b>49</b> |
| <b>APPENDIX B .....</b>   | <b>51</b> |
| <b>CURRICULUM VITAE .....</b>   | <b>52</b> |



## LIST OF FIGURES

|  |    |
|--|----|
| Figure 3.1 Counterspeech Treatment .....                             | 21 |
| Figure 3.2 Hate Speech Treatment .....                               | 22 |
| Figure 3.3 Code Scheme of Analyzing Open-ended Questions.....        | 28 |
| Figure 3.4 Examples of Hateful Comments Written by Participants..... | 28 |



## LIST OF TABLES

|   |    |
|---|----|
| Table 3.1 Summary Statistics of Demographic Characteristics and Dependent Variable..... | 25 |
| Table 3.2 Frequency of Demographic Characteristics.....                                 | 26 |
| Table 4.1 Percentage of Preexisting perceptions of Participants Toward Refugees.....    | 33 |



## LIST OF ABBREVIATIONS

AKP: Adalet ve Kalkınma Partisi (Justice and Development Party)

AMER: Association for Monitoring Equal Rights

CHP: Cumhuriyet Halk Partisi (Republican People's Party)

CoE: Council of Europe

CSOs: Civil Society Organizations

HDF: Hrant Dink Foundation

HDP: Hakların Demokratik Partisi (People's Democratic Party)

HRA: Human Rights Association

ICTs: Information and Communication Technologies

LGBTİ+: Lesbian, gay, bisexual, transgender, and intersex

RG: Reconquista Germanica

RI: Reconquista Internet

STK: Sivil Toplum Kuruluşu

TüSES: Türkiye Sosyal Ekonomik ve Siyasal Araştırmalar Vakfı (Turkish Economic Social and Political Research Foundation)

UN: United Nations

UNESCO: United Nations Educational, Scientific and Cultural Organization

UNHCR: United Nations High Commissioner for Refugees

## 1. INTRODUCTION

While social media platforms offer many benefits, e.g., limitless and low-cost communication all over the world, generating income, endless entertainment and sharing ideas and thought worldwide; it is also a sphere that allows people to disseminate their hateful ideas and beliefs. With the advancement of digital technologies, access to the internet and these platforms have become easier. Yet, the ease of sharing hateful thoughts on social media platforms and the visibility of them have also rapidly increasing.

In 2015 the Council of Europe's (CoE) Anti-Racism Commission described online hate speech as a phenomenon and a growing problem in many countries (Bakalis 2015 11). At the same time, reports published by Facebook, Instagram, Twitter, and YouTube also indicate a rise in hate speech on social media platforms. Between January and March 2021, 85,247 videos on Youtube, 25.2 million pieces of content on Facebook and 6.3 million pieces of content on Instagram were removed or flagged for violating these social media platforms' hate speech policies (United Nations Educational, Scientific and Cultural Organization [UNESCO] 2021, 6). It is seen that there is a 427% increase in the number of removed contents by Instagram between 2019 and 2020. Similarly, while YouTube removed 2,144,667 contents in 2019, this rate increased to 5,048,897 in 2020 (Reboot 2020). A similar situation is seen in Twitter's transparency report. Between July and December 2020, the number of accounts removed from Twitter for violations of its hateful conduct policy had increased by 77%, from 635,415 to 1,126,990, the highest number in recorded history (Twitter 2021). In Turkey, 36,578 pieces of content was removed in the first half of 2021 (Twitter Transparency Report of Turkey June 2021, 4), while in the second half, 43,430 pieces of content was removed by Twitter on the grounds that it was against its hate speech policy (Twitter Transparency Report of Turkey January 2021, 6).

There is no definition for hate speech that is internationally accepted by all, but it has mostly manifested itself in different forms through a variety of expressions, e.g., calls

for discrimination or exclusion, using harmful stereotypes, incitement to violence or hostility, insulting, misinformation, demonization, dehumanization of a group of people based on their race, ethnicity, gender, sexual orientation, religious grounds, age, national origin, gender identity, caste or disability and immigration status (Allan 2013; Garland et al. 2020, 3; Twitter 2021; YouTube 2019; Weber 2009). According to the existing literature, online hate speech is frequently directed toward marginalized groups (Kim, Sim, and Cho 2022, 3), e.g., women (Saha et al. 2018), the Lesbian, gay, bisexual, transgender, and intersex (LGBTI+) community (Strand and Svensson 2021), immigrants and refugees (Dinar et al. 2016). Hateful ideas against these groups can be disseminated through comments, memes (images and videos), and posted publicly or/via closed messaging applications on online platforms (Miškolci et al. 3).

Particularly after the Syrian civil war, the number of asylum seekers and refugees<sup>1</sup> increased rapidly worldwide (The Refugee Project 2020). As a result of that, these groups are perceived as an economic burden, a security risk and a threat by host countries (Pak and Elitsoy 2020, 581; Çoşkan, Erdugan and Oner-Ozkan 2022). Since Turkey hosts the highest number of refugees worldwide, with more than 3.7 million (United Nations High Commissioner for Refugees [UNHCR] 2021), negative attitudes towards refugees are encountered both offline and online (Müller and Schwarz 2021, 2132). In 2014, a group of young people who wanted the Syrians to leave Turkey organized on social media platforms under the "#Idon'tWantSyriansInMyCountry" for the first time and gathered to protest in Kahramanmaraş, Turkey. During the protest, the group attacked a car with a Syrian family and wanted to lynch them, and signs of some workplaces belonging to Syrians were also taken down (Hürriyet 2014; NTV 2014).

As exemplified above, the spread of hate speech content brings several negative consequences. These consequences not only end up in hate crime but also implicate psychological harm and social problems, e.g., depression, anxiety, drug abuse, polarization, extremist mobilization and radicalization (Chaudhary, Saxena and Meng

---

<sup>1</sup> In this thesis, refugees are identified based on the definition of The United Nations Refugee Agency (The UN Refugee Agency). "Refugees are people who have fled war, violence, conflict or persecution and have crossed an international border to find safety in another country." (The UN Refugee Agency, n.d.).



2019, 3-4; Cinelli et al. 2021; McDoom 2012). Therefore, social media companies, CSOs, governments and individual activists develop strategies to combat online hate speech and reduce its impacts, e.g., removal of hateful content, automated content moderation and censorship or banning an account. However, these strategies could generate different consequences, including damaging or abridging the freedom of speech, causing the spread of hate to other platforms instead of removing it and technical difficulties (Garland et al. 2020 3; MacAvaney et al. 2019). To curb these results, counterspeech is becoming more and more popular as a rising alternative strategy to combat online hate speech (Hangartner et al. 2021, 1). The thesis focuses on the potential of counterspeech in alleviating the consequences of hate speech online.

Counterspeech is a direct response given to hateful comments by online platform users (Benesh et al. 2017, 24). The purpose of the counterspeech is to curb the spread of hatred online, reduce its effects, as well as impact and change the users' behavior (Buerger 2020, 2). As with hate speech, counterspeech also comes in different forms, e.g., correcting misinformation in hateful messages, supporting victims, developing neutral comments, interacting (like, comments, resharing) with other counterspeech messages, and explaining the consequences of hate speech (Garland et al. 2020, 3).

Despite a growing number of groups and individual activists that use counterspeech to struggle against hatred online, e.g., #jagärhar (#iamhere), @pangazar (Flower Speech), @YesYoureRacist (Dangerous Speech Project n.d), there has been no such effort made to respond directly to hatred online in Turkey. Despite the increase of using this method around the world, the experimental evidence testing the effects of counterspeech as an intervention is quite limited. In the case of Turkey, most of the studies related to hatred online are based on discourse analysis (Erdoğan-Öztürk and Işık-Güler 2020; Erbayal-Filibeli and Ertuna 2020; Aslan 2018; Süllü-Duru and Yılmaz-Altuntaş 2019). In this sense, this study will be the first research in Turkey conducted on the effectiveness of counterspeech by using the online survey experiment. With this thesis, I aim to demonstrate whether counterspeech can be used as an effective method in Turkey to combat online hate speech by asking the following research question: Do participants

who are only exposed to empathy-based counterspeech comments write less hate speech than those who read only hateful comments against refugees on Twitter?



## 2. LITERATURE REVIEW

The primary purpose of this study is to understand whether the counterspeech method can be used as an effective strategy to combat online hate speech in Turkey. Recently, there have been a lot of studies worldwide on hatred online and how to best counter it. In Turkey, despite having conducted several studies with regards to online hate speech, there has been no systematic research into developing an effective method to combat this serious problem. This chapter will briefly give background information about offline and online hate speech, studies conducted on online hate speech in Turkey, and a general framework about counterspeech and the research done on this topic. Finally, I will list the objectives and the scope of the study along with the gaps that I aim to fill within the existing research literature.

### 2.1 Hate Speech

Even though several studies have conceptualized and identify hate speech in different disciplines, e.g., laws and social science (Baker 2008; Weber 2009; Sellars 2016; Brown 2017), there is still no universally accepted definition. One of the most common definitions for hate speech used by scholars was issued by the CoE of Ministers in 1997.

"The term "hate speech" shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin." (107).

Similarly, the United Nations (UN) does not give a specific definition for this term but characterize hate speech as communication that includes insulting or discriminatory discourse in written, spoken, or behavioral forms against a person or group based on their identities (2019). Although international organizations such as the UN and EoC define hate speech differently, reaching a consensus on the framework of the term is essential to understand the scope of hate speech and prevent and point out its negative consequences.

Hate speech emerges in many different contexts and categories. Binark and Çomu (2012) categorize hate speech under the six different contexts: “Political Hate Speech, hate speech against women, hate speech against foreigners and immigrants, sexual-identity based hate speech, religious belief and sect-based hate speech and hate speech against disabled people and diverse diseases”. The common feature of all these categories is directed toward outsiders groups.

Hrant Dink Foundation (HDF) is one of the organizations in Turkey that carry out its activities mostly on hate speech, issued a report in 2014. According to the report there are 4 categories of hateful discourse in Turkey: “Exaggeration / Attribution / Distortion, Blasphemy / Insult / Degradation, Enmity / War Discourse and Use of inherent identity as an element of hate or humiliation / Symbolization” (The Hate Speech and Discriminatory Discourse 2014, 11). The importance of this report and categories is that these categories were identified by scanning all national newspapers and almost 500 local newspapers and through systematic research carried out on this issue considering Turkey's language and cultural differences. Therefore, this report draws a route to understand the hate speech ecosystem of Turkey.

On the other hand, while trying to define and combat offline hate speech, and with the development of Information and Communication Technologies (ICTs), a new sphere has emerged in which hate speech can be spread: The internet. Particularly, after the growth of Web 2.0 technologies through the advent of social media platforms in the 2000s, the general picture of online hate speech has changed dramatically. As an internet-based application, social media platforms allow individuals to develop and share different pieces of content, e.g., photos, videos, and opinions (Cohen-Almagor 2015, 21). Although this seems like a unique feature within the scope of freedom of expression, it has also brought about a sphere where hate speech against subalterns is rapidly produced and spread. Stormfront, which can be called the first major hate group that uses white nationalism and other forms of radicalism, emerged quickly by using the internet as a popular tool in 1995 (De Gibert et al. 2018, 11). By 2020, this number had reached 838 in the USA (SPLC Southern Poverty Law, 2021).

At its simplest, online hate speech is a type of hate speech that takes place in any ICTs (Rudnicki and Steiger 2020, 7). Unlike offline hate speech, it can be disseminated in different forms of communication, e.g., image, video, written or voice record (McGonagle 2013). Furthermore, some features of the online, e.g., anonymity, invisibility, organizing domestic or international hatred communities quickly and reaching more people with instant publishing, can have different impacts which contrast to the offline environment (Brown 2018). In other words, hateful thoughts and beliefs can be developed and go viral quickly online. Therefore, they can reach numerous people in seconds and cause repeated victimization (Leonhard et al. 2018, as cited in Obermaier et al. 2021, 560).

Similarly, as with offline hate speech, online hate speech can also manifest in different forms: Intolerance towards other groups, negative ideas toward targeted victims, spreading misinformation related to victims, acting as a member of a targeted group, intellectual forms of hate speech and "trusted information" (Gelashvili 2018, 44-45). While intolerance, negative ideas and spreading misinformation are more traditional ways to generate hate speech, other forms have emerged with the social media platforms.

The purpose of the intellectual forms is not only to spread hatred online but also to legitimize their messages (Klein 2012, 437-438). Some racist websites like "Institute for Historical Review", which disseminates false information on Holocaust, try to appeal to their visitors by creating a false impression of a scientific approach. The said website, for instance, refers their resources with an academic jargon, with which they aim to legitimize the anti-semitic information they circulate.

With the development of social media platforms, people have been exposed to an increase in content and information. This has led to their information trust threshold being increased (Klein 2012, 440). Therefore, it has disseminated through social media platforms without considering the accuracy of the information given. For example, a journalist who wrote a column for a national newspaper claimed that Syrians received salaries from the state. The columnist cited a thesis written at Yavuz Selim University

as the source of this information. However, there is no such university in Turkey. This article has been shared repeatedly on social media platforms, targeting Syrians (Teyit 2019b).

Online hate speech can disguise itself as support in some cases. A webpage, which usually supports violence or terrorism, may seem to be promoting an agenda, while this webpage undermines this very cause by spreading information that could provoke a public rage (Jakubowicz et al. 2017, 58).

It is not possible to consider hate speech independently of the emotion of hate. Several studies argue that hate speech against an out-group is enhanced due to triggering or increasing the feeling of hatred because of different sorts of information (Bahador et al. 2021). Waltman and Mattheis (2017) associate hate with a lack of empathy which can hurt and cause violence against a group or an individual. For example, research conducted on a group of bystanders showed that individuals who had experienced hate speech were more likely to act as bystanders, as opposed to others who have not been victimized. This was due to a feeling of understanding, fear, felt towards them (Henson, Fisher and Reynolds 2020; Wachs and Wright 2017). Understanding the emotion of hate and its causes is critical in determining the direction of methods developed to combat hate speech. In particular, it is fundamental to understand the reason for this emotion and for the strategies used in one-to-one communication with haters, e.g., counterspeech.

While some studies state that for a piece of content to be considered hate speech, it should not be directed to individuals, and their characters (Hawdon, Oksanen and Rasanen 2017), another group of researchers argue that online hate speech can target both individuals and groups; however, the consequences might differ (Latour et al. 2017, 33). For example, when online hate speech is directed towards a group, one of the aspects of it arises in the form of dehumanization and demonization, which then leads its victims being described as inhuman. This in turn causes all members of that group to be labeled the same just by association, even if it was only a minority within it who were responsible for the supposedly negative actions (Bahador 2021).

Regardless of whether the hatred produced is directed against an individual or a group, it is seen in the studies that it has negative effects both on the communities and on the individual. For instance, being exposed to insulting or discriminatory language affect individuals' mental health and cause depression or anxiety (Chaudhary, Saxena and Meng 2019). According to Cinelli et al. (2021), there is a direct relationship between hate speech and polarization, as online platform users are more prone to express hateful thoughts against an out-group. Furthermore, a significant number of studies have shown that the consequences of online hate speech almost always have a real-life reflection (Jakubowicz et al. 2017). According to a report published by Human Rights Association (HRA) between 2010 and 2020, of the 280 people who had been racially attacked, 15 people were murdered, 3 were Syrian children, and a further 1097 were injured in this period because of these attacks. The report also reveals that hate attacks against Syrians, one of the groups most exposed to hate speech, have increased (Human Rights Association 2020, 5).

To prevent potential negative consequences of hate speech, CSOs, international and local institutions, governments, social network services and scholars from different disciplines have developed strategies. However, it is challenging because while combating online hate speech it is also important to protect ethical issues such as freedom of speech or human rights. Commonly used strategies, e.g., removal of content automatically, banning an account or censorship, for this purpose, are carried out by technology companies or social network services; therefore, much criticism emerges regarding such transparency and accountability of the process carried out during the censorship or banning of accounts (Laub 2019). Another problem is that these platforms use multitude of languages, making it difficult for automatic detectors to catch hate speech.

## **2.2 Online Hate Speech in Turkey**

Nine countries around the world accept online hate speech as a crime (Kaos GL 2020). Yet, Turkey is not one of those countries (IHD 2020, 2). Furthermore, although on

October 1, 2020, a law called On Regulation of Publications on The Internet and Combating Crimes Committed by Means of Such Publication ("The Internet Law") entered into force in Turkey (İnceoğlu, Sözeri and Erbaysal-Filibeli 2021, 7), it does not contain any articles on preventing the spread of hate speech on the internet (Resmi Gazete 2020).

In addition, not having a legal system to prevent the impacts of online hate speech, CSOs, individual activists and scholars hardly ever carry out activities or develop strategies to reduce online hate speech's effects and combat it. Even though some institutions, e.g., HDF, Kaos GL, Association for Monitoring Equal Rights (AMER), implement activities on online hate speech and its consequences, they are mainly focused on developing an archive on this issue as well as monitoring of hate speech against specific groups. Therefore, it is difficult to argue that activities carried out in civil society differ from the existing academic research on this issue. Although some of those studies and activities develop various suggestions to prevent the spread of online hate speech (Binark and Çomur 2012), unfortunately, to my knowledge, no research has analyzed the effectiveness of strategies to combat online hate speech in Turkey.

In Turkey as well as over the world, online hate speech emerges in various forms of expression based on people's identities, e.g., ageism, sexual orientation, and ethnic identity. Even though these expressions are directed toward different groups, they often appear in similar forms, e.g., insulting, dehumanization, marginalizing. To examine the genre of expressions are directed to aged people, Akgül (2020) conducted research on one of the popular platforms in Turkey, Ekşi Sözlük, where people can write anonymously. After analyzing 1794 comments under the "Curfew over 65+ years old" title, which opened after the curfew applied for individuals aged 65+ during the Covid-19 outbreak, Akgül found that 7.3 % of the comments under this title had hateful expressions. Moreover, the researcher identified frequently used hateful expressions toward this group: swearing, insults, humiliation and marginalizing. The LGBTI+ community is one of the groups receiving an inordinate targeting of online hate speech (Özatalay and Doğuş 2018, 18). Dondurucu (2018a) found that in İnci Sözlük, hateful comments based on sexual identity and orientation were produced collectively and in a



participatory manner under “gay,” “lesbian,” “bisexual,” and “transgender” titles. At the same time, the researcher identified the categories of hate speech directed toward the LGBTI+ community. 39.1 % of comments indicate homosexuality as a figure of humor, comedy and entertainment. While 39.1 % of comments indicate homosexuality as a figure of humor, comedy, and entertainment, 29% of collected comments describe being homosexual as a social deviation, and finally, 7.25% characterize homosexuality as a psychological illness (Dondurucu 2018a 1393). In addition to that, Dondurucu concluded that the users produced intense hate speech against the LGBTI+ community as well as representing this group in a stereotypical way in the new media environment due to their sexual orientation. In another study, Dondurucu (2018b) analyzed the comments shared on Twitter under the hashtag #homosexuality. It was found that 86.4% of these tweets contain negative comments toward this group (527). At the same time, the most common forms of hate speech under #homosexuality hashtag emerged as marginalization and discriminatory expressions. According to the report published by Kaos GL (2020), 2028 discriminatory discourses toward LGBTI+s were developed in 2020 in the news, interviews, and op-ed pieces. In the pieces of content, homosexuality was shown as “perversion” in 44%, “disease,” in 37%, and "sin" in 41% of them (Kaos GL 2021).

Although there are refugees and asylum seekers in Turkey from Iraq, Afghanistan and other countries (Asylum Information Database 2021), research related to online hate speech and refugees has been primarily focused on Syrian refugees. This is due to the Syrian refugee situation being a constantly perceived problem to both politicians and traditional media outlets alike (Özkaynak and Doğuş 2018, 4). As mentioned before, most academic studies related to online hate speech are based on discourse analysis. In particular, online hate speech research conducted on refugees is formed within the framework of Syrian refugees and discourse analysis. Kurt (2019) analyzed hateful comments under the five most-watched videos related to Syrian refugees on Youtube. The researcher found that Syrians are often exposed to insults, slander, racism, and hostility. Moreover, Kurt discovered that Syrian refugees are one of the largest recipients of online hate speech. Similarly, Alikılıç, Gökaliçer and Alikılıç (2021) conducted research to determine what kind of discriminatory discourses are mostly used

toward Syrian refugees. To this end, 4217 tweets were analyzed under the hashtags #syrian #syrianrefugee #refugee #WeDon'tWantSyrianRefugeesInOurCountry and #Syrianout!. They found that the discourses against Syrian refugees were marginalizing, dehumanization and incitement to violence and contained hostility. Drawing on over 1500 tweets under #WeDon'tWantSyrianRefugeesInOurCountry hashtag on Twitter, Taşdelen (2020) found that Syrian are seen as invasive, greedy and immoral. Moreover, insults were frequently aimed towards this group. According to the above-mentioned literature, online hate speech toward outcast identities in Turkey manifest itself in different forms through a variety of expressions, e.g., humiliation, marginalization, incitement to violence, insulting, and dehumanization.

### **2.3 Counterspeech**

With the development of social media platforms, internet users encounter online hate speech more than ever. Therefore, researchers from a variety of different fields, social network services, and governments have attempted to develop strategies in order to tackle this problem and reduce its effects. Among these methods, counterspeech is a strategy for direct response to hate speech online (Benesh et al. 2017, 5). Counterspeech is also defined as a strategy developed based on disagreement and expressing counterview toward hateful and extreme thoughts and beliefs online (Barlett and Krasodomki-Jones 2015, 5). Benesch (2020) describes counterspeech as a method that does not infringe the right of freedom of expression as well as creates an online environment in order to reduce the negative impacts of hate speech (23). Benesch's counterspeech definition, emphasizing this strategy does not harm right of freedom of expression, cannot be considered independently from the history of counterspeech. First Amendment jurisprudence in the USA indicates that to combat hate speech and reduce its harmful effects, not excluding hate speech from areas where there are free debates, to struggle with this problem with more counterspeech (English 2021, 13). It is mainly linked with the Justice Brandeis's saying in 1927 during the case of *Whitney V. California* on the government's right to suppress dissent: "the solution to bad speech is more speech" (Cepollaro, Lepoutre and Simpson 2022, 2).

Counterspeech is categorized into two types: "organized counter-messaging campaigns and spontaneous-organic responses" (Benesh et al. 2016, 7). Counterspeech can be generated by different counter-speakers or/and groups from various social, cultural, and economic backgrounds, such as victims, bystanders, people who are not targets of hate speech, everyday citizens, authoritative roles, e.g., CSOs, governments, and institutions (Cepollaro, Lepoutre and Simpson 2022, 3). An example of a group work, Reconquista Internet, founded in Germany in 2018 by Jan Böhmermann to respond collectively to hateful content on Twitter produced by the far-right group Reconquista Germanica, today has more than 50,000 members (Keller and Askanius 2020, 544). There are also many individual activists producing counterspeech on their own. For example, since 2016, Journalist Hasan Kazim<sup>2</sup> has been responding to each xenophobic message he received from Twitter users. Iyad el-Baghdadi<sup>3</sup> generates counterspeech to respond to hateful content that he encounters on his personal Twitter account.

Another important point about this strategy is that the receiver of the counterspeech also varies, which is significant since it affects the counterspeech's power and effectiveness. The receiver may consist of victims, people who might support the victims and generate more counterspeech, individuals who have negative perceptions toward targeted groups or a combination of these (Saul 2021, 5).

Similar to online hate speech, counterspeech also manifests itself in various forms. The first one is factual counterspeech which mainly refers to a "deliberative discussion atmosphere." This atmosphere is developed to point to prejudice and misinformation they produce (Obermaier, Schmuck and Saleem 2021, 3; Garland, 2020, 3). The other strategy to counterspeech is to share personal experiences (Miškolci, Kovacova and Rigova 2018, 4); thus, empathy can be established with the target group. Besides these two strategies, there are also other forms, e.g., warning about the consequences of hate speech, interacting (Like, comments, resharing) with other counterspeech messages, and explaining the effects of hate speech (Buerger 2020 10; 2020, 10; Garland et al. 2020 3).

---

<sup>2</sup> <https://twitter.com/hasnainkazim>

<sup>3</sup> [https://twitter.com/iyad\\_elbaghdadi?ref\\_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor](https://twitter.com/iyad_elbaghdadi?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor)

The foremost aim of counterspeech is to change the beliefs and ideas of a person who spreads hatred online (Buerger 2021a, 2). However, this is not the primary objective for all organized groups using counterspeech. For example, one of the biggest organized counterspeech groups based in Sweden, #iamhere (#jagärhar), collectively responds to hateful comments, aiming to make their counter comments more visible and reach as broad an audience as possible. (Buerger 2021b). The group has approximately 74.000 members and has been countering hate on Facebook in 16 countries, such as France, Canada, Spain, and the Czech Republic since 2016 (5). The #iamhere hashtag seeks to increase the visibility of counterspeech while simultaneously increasing the impact of these comments through a set of rules such as writing civil comments or using a productive tone when replying.

### **2.3.1 Factors that affect the effectiveness of counterspeech**

It is no coincidence that groups like #iamhere consider engaging in a good dialogue while countering hatred online. Bartlett, Jamie and Krasodowski-Jones (2015) found that forms and tones of comments are vital in order to reach more people. For instance, comments with questions as well as posts including humorous and sarcastic counterspeech messages receive more integrations by comparison with all the other forms they studied. Similarly, Frenett and Mott (2015) analyzed how the tone of counterspeech caused a reaction, e.g., blocking or sending a message to counterspeakers, from Facebook users who write hateful comments. As a result, the authors found that the tone of the counter-message immensely affected the reaction of the Facebook users. For example, while they have never received a response to Antagonistic messages, casual and sentimental messages received more than 80% response (17).

Miškolci, Kováčová and Rigová (2018) analyzed more than 7000 comments written on Facebook on Rome-related topics to demonstrate how the identity of the Roma people living in Slovakia was built as well as the effect of counterspeech responses directly hate speakers. The authors found that counterspeech is not an effective strategy to

reduce online hate speech against Roma. However, they discovered that posting pro-Roma counterspeech motivates other audiences to get active. Garland et al. (2020) analyzed 180,000 tweets collected over four years, belonging to Reconquista Germanica (RG), a far-right troll group and Reconquista Internet (RI), the largest counterspeech group aims to counter RG's hateful messaging in Germany. They found that after RI was established, the frequency and volume of hate speech decreased. Moreover, showing support to counterspeech messages, e.g., likes, and comments, as well as making them more visible through organized counterspeech, reduces the number of hate speech.

To determine the factors that affect the effectiveness of counterspeech on Facebook, Schieb and Preuss (2016) conducted a computational simulation model. The authors found that counterspeech can have a significant impact on a given audience depending on the volume of the hateful comments. In other words, the effect of counterspeech increases as the volume of hate speech comments decreases. Cheng et al. (2017) also conducted an online experiment. The authors found that participants produce comments similar to the comments they are exposed to. For example, when someone is exposed to negative comments, the probability of posting negative comments or responses increases.

Since one of the main purposes of counterspeech is to change people's attitudes and behaviour while commenting, few studies have shown how counterspeech changes these factors. Han and Brazeal (2015) conducted an online experiment to find whether civil or uncivil political discourse affects how people discuss. The authors found that those exposed to civilised interpretations also apply this within their own interpretations. At the same time, it was seen that this group was more willing to participate in the discussions compared to the group that was exposed to uncivil comments. In order to test the effectiveness of one of the German organised counterspeech groups #ichbinhie users' comments ("I am here"), Friess, Ziegel, and Heinbach (2021) analysed more than 124.000 comments between November 01, 2017, and January 31, 2018, via quantitative content analysis. The purpose of the study was to answer the following questions: Are the quality of comments posted by #ichbinhie

members higher than the quality of comments written by non-members of the group? Does using a respectful and civil tone while developing counterspeech encourage others to comment similar way? They found that comments written by #ichbinhie are more considerate than those written by non-members, as well as #ichbinhie members affect the quality of discourse on Facebook. On the other hand, civil and deliberative counter messages stimulate more respectful and polite comments. A similar result emerged in the study conducted by Molina and Jennings (2018) to test whether making civilized comments affected how other participants commented on Facebook. The authors found that being exposed to civil comments and metacommunication stimulated others to participate in the discussion and write respectful comments. In other words, civil comments affect participants' commenting behaviour positively.

Hangartner et al. (2021) designed a field experiment based on three interventions: "empathy, warning of consequences, and humor" to reduce xenophobic and racist hate speech on Twitter. By analyzing more than 1000 tweets, the authors found that empathy-based counterspeech messages increase the deletion of previously produced hate speech. At the same time, it revealed that that kind of message reduces the occurrence of possible hate speech by containing xenophobia. Another online experiment was conducted by Benjumea and Winter (2018) to measure the effectiveness of interventions on the audiences: Counterspeech and deleting of hateful comments (Censorship). The authors tested whether that kind of intervention affected the audiences' behaviour while posting a subsequent comment in the same area. They found that the audience posted less hateful comments when they faced censorship; however, they did not find a similar result using counterspeech.

Most of the studies related to counterspeech strategy have been conducted in Western countries such as Sweden, the UK, and France. With this study, I aim to fill a gap in this literature by being the first study conducted in developing countries, as opposed to developed ones, that measure the effectiveness of counterspeech strategy in challenging hatred online in Turkey. The number of studies using the online survey method to measure the impact of counterspeech is quite limited- Kim, Sim and Cho (2022) conducted an online survey experiment with 1250 people in South Korea to understand

how gender and the popularity of counterspeech affect the reporting of hate speech produced around the #MeToo movement on YouTube. The authors found that YouTube users are more likely to report misogynist discourse if counterspeech comes from a female. Nevertheless, when a female counter speaker receives numerous “Likes”, it particularly affects male users' attitudes, who are less willing to report hate speech. To understand the determinants of participation in counterspeech using an online survey experience, Kunts et al. (2021) carried out research in which they tested whether the norms of citizenship solidarity encourage "online civil intervention" on a specifically designed online news site. The authors found that internet users' willingness to engage in counterspeech against hateful comments depends on the social group being attacked. Obermaier et al. (2021) also used an online survey experiment to understand how Muslims living in Germany react to online hate speech and how counterspeech generated by the majority, as well as different minority groups, influences Muslims' response. The authors observed that Islamophobic comments generated on Facebook are perceived as a threat by Muslims; therefore, their intention to produce counterspeech increases. At the same time, they found that the counterspeech generated by the majority or minority groups decreases the counterspeech behavior of Muslims. Unlike the three studies mentioned above, this study is not interested in the factors that affect counterspeech or the motivation to be a counter-speaker. With this study, I aim to fill the gap in the English literature by testing how exposure to only counterspeech comments affects participants' hate speech-generating behavior on Twitter by using an online survey experiment. This thesis is also the first study on refugees conducted directly using counterspeech and online survey experiments. Overall, this thesis seeks to answer the following research question:

Research Question: Do participants who are only exposed to empathy-based counterspeech comments write less hate speech than those who read only hateful comments against refugees on Twitter?

### **3. METHODOLOGY**

The research question in this study asks is whether counterspeech can be used as a strategy to combat online hate speech in Turkey. To answer this question, an online survey experiment with a sample of 181 people over the age of 18 who can speak Turkish was conducted.

#### **3.1. Survey Experiment**

To measure whether counterspeech messaging makes a difference in the hateful comment posting behavior of social media users, I conducted an online nonprobability survey experiment in Turkey. Schnabel (2021) describes survey experiments as "an experiment conducted on a survey (32)." A survey experiment is effective as it allows insight into universality and causality. At the same time, by giving anonymity to the respondents, survey experiments for measurement offer a more honest and robust response to sensitive topics (Diaz, Grady and Kuklinski 2020). Furthermore, in a country like Turkey, where social media posts are often cited as criminal evidence (Human Rights Association 2021), anonymity is key for ensuring response honesty hence crucial for the right interpretation of experimental results. Therefore, using a survey experiment in this thesis allowed me to evaluate whether the participants who were accidentally exposed to the opposite discourse produced expressions that could be defined as hate speech. These include humiliation, insult, incitement, and dehumanization.

#### **3.2 Design of the Experiment**

To collect data to answer the research questions posed for this study, a 2 X 1 between-subjects experimental design was carried out. Participants were randomly assigned to one of the two conditions.



After the informed consent procedure, all participants were exposed to two original news stories about refugees posted on two popular newspapers' Twitter pages: "Banana Eating Videos" and "Wall on Iran Border." The two pieces of news were selected based on both showing the impacts of online hate speech on real-life circumstances, with the first questioning the access of basic needs of refugees by dehumanizing them, and the later drawing attention to a different refugee group other than Syrians. To design realistic Twitter threads visuals, I used the White Bird application, which allows for generating fake tweet threads. While creating the Twitter threads visuals, I also utilized two online platforms: one as a random name generator ([behindthename.com](http://behindthename.com)) and the other for fake faces ([thispersondoesnotexist.com](http://thispersondoesnotexist.com)). Moreover, since political opinion is one of the essential factors affecting trust in the media and news (Teyit 2019, 32), in this study, the logos and names of the Twitter account from where the news was taken were blurred in order to prevent any bias.

One group saw both news posts with only hateful comments, while the other group was shown the posts with only counterspeech comments; and each participant was asked to respond with their own comments under the posts. Participants first saw the news "Banana Eating Videos" with the original caption and picture taken from the Twitter account of the news source. Then, the respondents read the "Wall on Iran Border" news, which was similarly prepared. The hateful sample included five comments in the form of five different hate speeches: Dehumanization, insulting, misinformation, incitement to hostility and exclusion. Moreover, in the hateful samples, refugees were also demonstrated as an imposition ("I don't want refugees in my country. I hope they are all deported."). To create hateful comments under the shown news, I analyzed more than 500 comments related to refugees on Twitter, Ekşi Sözlük, Facebook and YouTube. Taking these comments into consideration, I generated new hateful comments under the presented news. In the same manner as the hateful condition, the counterspeech sample consisted of five comments in two different forms of counterspeech: Sharing personal experiences and supporting victims using empathy-based content ("72 % of young people in Turkey are looking for a way to go European countries. I wonder if Europe will consider a banana as too much for our own children."). While creating counterspeech comments, I had originally planned to apply the same process as with the

hateful condition, but there were too few examples that could be identified as counterspeech. Therefore, I utilized a digital platform called Toolkit for Human Rights Speech<sup>4</sup> to protect human rights and democratic principles while writing counterspeech comments. All comments for both cases were under 280 characters, considering Twitter's character limit feature. Please see **Figure 3.1** for the counterspeech treatment and in the **Figure 3.2** hate speech treatment.




---

<sup>4</sup> The Toolkit for human rights speech can be reached via the following address, <https://pjp-eu.coe.int/en/web/human-rights-speech>.

### Banana Eating Video

İzmir'deki bir vatandaşın sokak röportajında "Ben muz yiyemiyorum, onlar kilolarca muz alıyor" sözlerinin ardından Suriyeliler'in sosyal medya üzerinden paylaşmaya başladığı "muz yeme" videoları sonrasında 8 Suriyeli gözaltına alındı.



15:53 · 30 Kas 21 · TweetDeck

17 Retweets 8 Quote Tweets 172 Likes

**Belma Badem** @belmabadem · 4s  
Replying to @  
Çok üzüldüm bu habere... Ne olmuş yani? Tıpkı bizler gibi bu insanlar da bir şeyleri protesto etmiş. Gerçek adaletin gelmesi için hepimiz için adaleti talep etmeliyiz...

**Ahmed Uzun** @uzun5784 · 4s  
Replying to @  
Bir grubun yaptığı bir şey yüzünden tüm mültecileri suçlamamalıyız. Bir gün biz de mülteci olabiliriz, ülkemizi terk etmek zorunda kalabiliriz. Ve o gün bizim yediğimiz lokmalar da birinin gözüne batar


**Enes Küçük** @littleenes · 4s  
Replying to @  
Politikacılar ve devletler bu insanları canavarlaştırdı hepimizin gözünde... Oysa onlar da bizim gibi insan ve insanca yaşamayı hak ediyorlar

**Sevil Barış** @barissevil · 3s  
Replying to @  
Çaresiz insanlara yardım etmeye karşısanız böyle yönetilmedi de hak ediyorsunuz demektir. Suriyeliler neden muz yiyor diye değil biz neden yiyemiyoruz diye sorun önce

**Koray Özge** @kryozge72 · 3s  
Replying to @  
Türkiye'deki gençlerin %72'si Avrupa'ya gitmek istiyormuş. Düşünüyorum acaba Avrupa da bizim evlatlarımızı bir muzdu çok mu görececek

### Wall on Iran Border

Van Valisi Bilmez: Sınırdan yasadışı geçişleri engellemek için 295 kilometrelik İran sınırının tamamına duvar örülecek



08:00 · 27 Tem 21 · TweetDeck

22 Retweets 47 Quote Tweets 203 Likes

**Kismet Belgin** @kismetblgn80 · 2s  
Replying to @  
Bu insanların çoğu daha iyi bir hayat kurabilmek için savaşın parçaladığı ülkelerden kaçmaya çalışan insanlar... Kimin ne zaman mülteci olacağı belli olmaz. Biraz empati ve merhamet

**Abdülkerim Binici** @akerimbinici · 2s  
Replying to @  
Savaştan kaçıp ülkemize geldikleri için bu insanları suçlayacağınıza savaşa neden olanları suçlasaydık belki bugün bu halde olmazdık. Bu gördüğümüz Türkiye'den kaçmaya başlayacağız bakalım o zaman Avrupa ne yapacak bize

**Nedim Bardakçı** @nedimbardakci · 2s  
Replying to @  
Mültecilerden bahsederken onların da anası, babası, kardeşi, çocuğu olduğunu unutmayın! Eğer duvarları aşip gelyorlarsa bir nedeni vardır. Kimse sevdiklerini arkada bırakmak istemez

**Lale Değirmenci** @lalelidedgi · 2s  
Replying to @  
Lütfen Google'a mülteci kampı yazın ve karşınıza çıkan görsellere bakın... Ülkelerini terk ettiklerinde en iyi ihtimal böyle yerlerde yaşayacaklarını biliyorlar. Kimse zor durumda olmasa böyle bir hayatın içine gelmek istemez

**Ramazan Hurşit** @hursit96 · 1s  
Replying to @  
Sanırım bunun nasıl bir çaresizlik olduğu anlaşılıyor. Düşünsenize kendi ülkenizden kaçmak için hareket eden bir uçağın kanadına biniyorsunuz. Allah kimseyi böyle çaresizlik içerisinde bırakmasın

Figure 3.1: Counterspeech Treatment

### Banana Eating Video

### Wall on Iran Border

**Banana Eating Video**

İzmir'deki bir vatandaşın sokak röportajında "Ben muz yiyemiyorum, onlar kilolarca muz alıyor" sözlerinin ardından Suriyeliler'in sosyal medya üzerinden paylaşmaya başladığı "muz yeme" videoları sonrasında 8 Suriyeli gözaltına alındı.

15:53 · 30 Kas 21 · TweetDeck

17 Retweets 8 Quote Tweets 172 Likes

**Belma Badem** @belmabadem · 4s  
Replying to @  
Ülkemde mülteci istemiyorum. Umarım hepsi sınır dışı edilir

**Ahmed Uzun** @uzun5784 · 4s  
Replying to @  
Daha dün ağlayarak sınırdan geçiyordunuz. Şimdi de bizimle dalga mı geçiyorsunuz!! NANKÖRLER

**Enes Küçük** @littleenes · 4s  
Replying to @  
Ben bunlarla aynı ülkede yaşamak zorunda mıyım kadesim... YALLAH kendi ülkenize

**Sevil Barış** @barissevil · 3s  
Replying to @  
Kendi ülkemizde mülteci olduk. Mülteci biziz

**Koray Özge** @kryozge72 · 3s  
Replying to @  
Vergilerimizle bunlar lüks içinde yaşasın, biz muza bile hasret kalalım...

**Wall on Iran Border**

Van Valisi Bilmez: Sınırdan yasadışı geçişleri engellemek için 295 kilometrelik İran sınırının tamamına duvar örülecek

08:00 · 27 Tem 21 · TweetDeck

22 Retweets 47 Quote Tweets 203 Likes

**Kismet Belgin** @kismetblgn80 · 2s  
Replying to @  
Erken oldu!! Daha Afganistan'ın diğer yansı gelmedii! Onlar da gelsin sonra yükseltiriz duvarı..

**Abdülkerim Binici** @akerimbinici · 2s  
Replying to @  
Allah aşkına şu duvarların boyunu 2 metre yapmayın. İzlediğim görüntüler gerçekse üzerinden atlamak çocuk oyuncağı

**Nedim Bardakçı** @nedimbardakci · 2s  
Replying to @  
DUVAR ÇÖZÜM DEĞİL!!! Adamlar giden uçağın kanadına atlıyor. Kalıcı çözüm lazım bize

**Lale Değirmenci** @lalelidegi · 2s  
Replying to @  
Adam 1000 km yol yürüyör sınıra ulaşmak için. Duvar mı onu durduracak. Onlar da gelsin sonuçta Avrupa'nin mülteci bekçisiyiz

**Ramazan Hırşit** @hursit96 · 1s  
Replying to @  
Mülteciler gelmesin demiyorum ama şehir içlerinden kamplara götürülmeleri gerekiyor. En büyük hatamız onları şehirlere almak oldu..

Figure 3.2 Hate Speech Treatment

After participants were exposed to the news, I asked each participant the dependent variable question, concerning what they would post under these tweet threads, after reading pieces of the news and written comments; 51.9 % of the participants were

randomly assigned the hateful condition, while 48.1 % were given the counterspeech condition. After the exposure to the experimental stimuli, for the purpose of controlling the participants' characteristics, I collected sociodemographic data, e.g., gender, age, education level, ethnic identity, employment status and party voted for in the most recent elections. Lastly, in order to access participants' preexisting attitudes towards refugees, I used a modified eight-item version of the following scales: The refugee labeling scale (Önder, 2020) and the shorter version of the attitudes towards refugees scale (Doğan et al., 2017).

### **3.3. Experimental Stimuli**

Since hate speech on Twitter is spreading faster and deeper in Turkey compared to other online platforms (Yıldız 2018), I chose Twitter as the research platform. Another significant reason for selecting Twitter as a research site is that compared to other social media platforms, e.g., Instagram, Facebook, and TikTok, the characteristics of this platform have the potential power to manage individuals' perceptions (Kınay and Atalay 2021, 59). Furthermore, Twitter is used as an important source of news in Turkey (Polat, Dilmen and Sütçü 2021). The last reason was that the majority of the studies conducted in Turkey focused on discourse analysis of hate speech against refugees on Twitter (Taşdelen 2020; Yıldız 2018). Therefore, it offers a significant advantage in understanding the content of the discourses produced on Twitter for refugees.

Refugees are one of the groups most exposed to hate speech (Gelashvili 2018, 44). This situation is similar in Turkey. According to a report published by AMER (2018), discriminatory language is commonly aimed against three groups in Turkey: Kurds, LGBTI+s and refugees. However, a point that separates refugees from these two groups is that while tolerate discrimination towards Kurds and LGBTI+ in all areas of life is more highly than the average, this tolerance is even higher towards discrimination against refugees. (25). For this reason, while creating Twitter threads to be used as the experimental stimuli of the survey experiment, I decided to create two news stories about refugees based on actual news stories.

Two groups were exposed to the same news pieces of content but were provided with different written comments: hate speech and counterspeech - the participants were randomly assigned news content concerning refugees using the Qualtrics tool without any additional intervention.

One of these threads includes a newspaper article with the title “Banana Eating Videos,” in which Syrian people shared their videos on social media platforms eating bananas in protest after a citizen during a street interview claimed that “I can’t afford bananas. They (Syrian people) buy kilos of them.” (Bianet 2021). After the videos were posted, the hashtag #Idon’tWantSyriansInMyCountry became a trending topic on Turkey’s Twitter again. As a result of being targeted by numerous people on Twitter, 8 Syrians who shared those videos were detained for deportation.

The other Twitter thread was published with "Wall on Iran Border" (Hürriyet Daily News 2021). As a result of the Taliban's re-takeover of power in Afghanistan in the summer of 2021, many videos of Afghan immigrants entering Turkey were shared on social media platforms. Immediately after these videos, under the hashtags #Idon'tWantRefugeessInMyCountry and #Wedon'tWantAfgansInOurCountry, pieces of content with discriminatory, insulting, misinformed, and threatening comments and memes against refugees were spread on social media platforms. As a result of the increasing reactions, the Turkish Interior Minister announced that a wall will be built on the Iranian border (Hürriyet Daily News 2021). Both selected news articles as a sample are essential in understanding the consequences of hate speech directed at a group in an organized manner on social media.

### **3.4. Sample**

The online survey experiment was conducted between January 20 and March 18, 2022. With the help of the Qualtrics tool, I reached people who speak Turkish and are over 18 years old. To reach more people, Facebook and Instagram advertisements were placed twice, between February 4 and 8, and 12 -16 March. 2022. As compensation, two of the participants would be provided with a 100 TRY award. The total number of clicks on

the advertisements was 7,953 people. The total number of participants who filled in the survey was 202, but 21 of those responses did not fulfil the technical criteria, such as incomplete survey responses and/or not approving of the participant consent form, so there were 181 remaining participants.

The responder age range was grouped into four categories: (1) 18-29, (2) 30-44, (3), 45-64 and (4) 65+ (Age groups,  $M = 2.29$ ,  $SD = 0.808$ ). In each experimental group, the age group with the highest number of participants is the 2nd category, which is compatible with the age range distribution of the population in Turkey (TÜİK, 2022). Of the total participants, 44.8 % were male, 68.5% self-identified as Turks, 19.9 % as Kurds, and 11.6% as others (e.g., Armenian, Arab, Greek). The demographic characteristics of the sample is presented in **Table 3.1**.

**Table 3.1**

*Summary Statistics of Demographic Characteristics and Dependent Variable*

| Variables                                    | Obs. | Mean | Std. Dev. | Min. | Max. |
|--|------|------|-----------|------|------|
| <b>Dependent Variable</b>                    |      |      |           |      |      |
| Attitude toward the comments                 | 181  | 0,89 | 0,909     | 0    | 2    |
| Comments under "Banana Eating Videos" news + | 181  | 0,44 | 0,497     | 0    | 1    |
| Comments under "Wall on Iran Border" news +  | 181  | 0,47 | 0,5       | 0    | 1    |
| <b>Respondent Controls</b>                   |      |      |           |      |      |
| Attitude toward refugees                     | 181  | 3,32 | 1,06      | 1    | 5    |
| Gender                                       | 181  | 1,5  | 0,501     | 1    | 2    |
| Age  | 181  | 2,28 | 0,812     | 1    | 4    |
| Education level                              | 181  | 2,92 | 0,897     | 1    | 5    |
| Party identification                         | 181  | 3,46 | 2,067     | 1    | 7    |
| Ethnic identity                              | 181  | 1,98 | 0,683     | 1    | 4    |

*Notes.* Respondents' demographic characteristics were grouped into different categories. Please see more information for the categories in Appendix B.

Other than the gender and age distributions in the sample, there were incongruities between the sample and the demographic characteristics of Turkey. For example, the number of participants who voted for People's Democratic Party (Hakların Demokratik Partisi - HDP) and Republican People's Party (Cumhuriyet Halk Partisi - CHP) was higher than Justice and Development Party (Adalet ve Kalkınma Partisi - AKP). Please see the detailed demographic characteristics of participants in **Table 3.2**.

**Table 3.2***Frequency of Demographic Characteristics*

| Variables                                 | Frequency | Percentage |
|---|-----------|------------|
| Gender                                    |           |            |
| Women                                     | 91        | 50,30 %    |
| Men                                       | 90        | 49,70 %    |
| Age                                       |           |            |
| 18-29                                     | 30        | 16,60 %    |
| 30-44                                     | 81        | 44,80 %    |
| 45-64                                     | 59        | 32,60 %    |
| 65+                                       | 11        | 6,10 %     |
| Education Status                          |           |            |
| Graduated from secondary school and below | 11        | 6,10 %     |
| High school graduate                      | 37        | 20,40 %    |
| University graduate                       | 98        | 54,10 %    |
| Master's degree graduate                  | 25        | 13,80 %    |
| PhD graduate                              | 10        | 5,50 %     |
| Ethnic identity                           |           |            |
| Turks                                     | 127       | 70,20 %    |
| Kurds                                     | 34        | 18,80 %    |
| Unspecified                               | 10        | 5,50 %     |
| Other                                     | 10        | 5,50 %     |
| Voted Party                               |           |            |
| AKP                                       | 16        | 8,80 %     |
| CHP                                       | 68        | 37,60 %    |
| HDP                                       | 42        | 23,20 %    |
| MHP                                       | 3         | 1,70 %     |
| İyi Party                                 | 8         | 4,40 %     |
| Other                                     | 7         | 3,90 %     |
| Prefered not to say                       | 37        | 20,40 %    |

**3.5. Variables****3.5.1. Preexisting attitudes toward refugees**

To analyze participants' sentiments towards refugees, I adapted the "refugee labelling scale" (Önder, 2020) and the shorter version of the "attitudes towards refugees scale" (Doğan et al., 2017) to use a five-point Likert-type scale (1=strongly disagree to 5=strongly agree) with eight items. Example items are, "Turkey hosts the largest number of refugees in the world. Therefore, it should not take in any more refugees, and the borders should be closed." or "Refugees living in Turkey work and pay taxes. Therefore, they should have equal rights as Turkish citizens." This variable was used as a control variable during the analyzing phase. While the maximum mean score is 5, the minimum one is 1, so all eight items were scored into a mean index ( $M = 3.32$ ,  $SD =$



1.06) and its internal consistency was found to be acceptable. (Cronbach's  $\alpha = 0.92$ ). Accordingly, higher scores indicate that respondents have a more negative opinions towards refugees.

### **3.5.2. Dependent variable**

After exposure to the stimuli, participants were asked to write their comments with a minimum of 15 characters under each piece of news. It was the only section in the data collection phase where responses were collected with open-ended questions. This study defines a comment as an individual's idea and thought (Grabe et al., 2012). Single comments made under each news were considered as the unit of analysis. While the number of comments participants added under the hate speech condition was 188, the counterspeech conditions received a total of 176 comments. This sampling yielded 128 hateful comments, providing a total analysis sample of 364 comments. To identify participants' comments as hateful or not, a coding system was developed and used as an instrument to measure the score of the comments. These categories are mainly derived from the research literature on identifying, monitoring and mapping online hate speech (Gagliardone 2014; Kennedy et al. 2018; Article 19 2015; Council of Europe nd.)

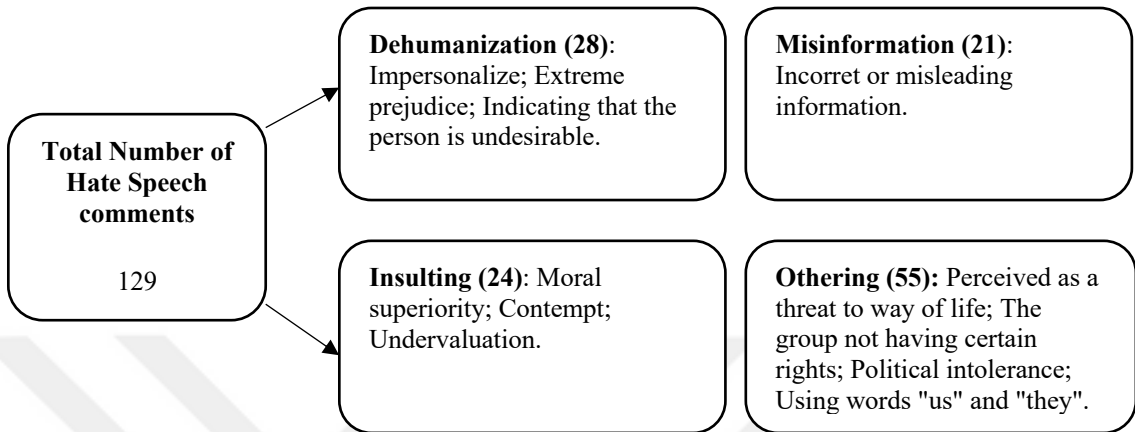
Identifying a case of hate speech is quite challenging since there is no universal definition (UNESCO 2021, 3). Furthermore, the fact that hate speech has numerous closely similar forms of expression makes it also difficult to identify (Quinn 2019, 6). In addition to these challenges, the lack of an accepted methodology for identifying hate speech causes it to be difficult to define what constitutes hate speech (Council of Europe nd.) Hate speech is based on prejudiced, harmful and offensive expression about the identities of a group or person (Article 19 2015, 9). In other words, hate speech is a negative generalization against a person or group due to their identities (Hrant Dink Foundation 2019, 15).

In this thesis, to classify hateful comments, four categories were determined taking into consideration defining identifiers of previous studies: Insulting, dehumanization, misinformation and othering. One comment could contain more than one different form

of hate speech. Below in **Figure 3.3** is the code scheme used to analyze the open-ended answers.

**Figure 3.3**

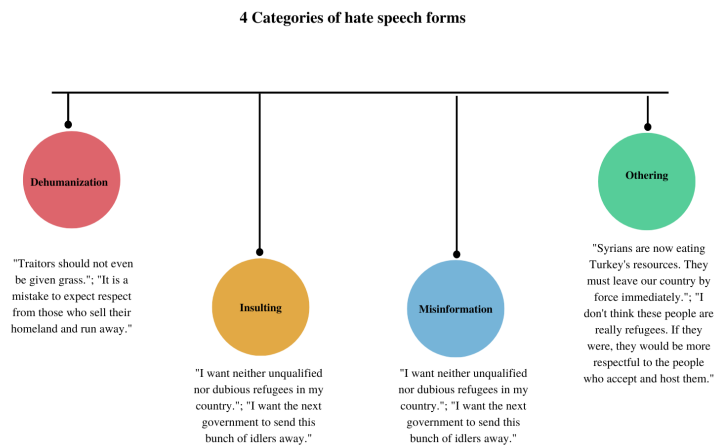
*Code Scheme of Analyzing Open-ended Questions*



To give more details how the comments were coded, I would share the following example: “For me, a refugee is a lowlife (Güruh) who needs to be helped, who has taken refuge in your country for some reason. However, this group (Bu kesim) is not in the same position for me. Of course, I don’t want people with this mentality in my country either. “This comment has three forms of hate speech: “lowlife” =insulting, “this group” =othering and dehumanization since it refers refugees as an undesirable group of people. Please see **Figure 3.4** for more examples of each category below:

**Figure 3.4**

*Examples of Hateful Comments Written by Participants*



Once a comment is identified as hateful, to measure the score of a participant's comments, an instrument was designed and scored as follows: (a) 1= including a hate form or forms and (b) 0=no hate form, to create a continuous variable. While the highest score a participant could get is 2, the lowest one is 0 ( $M = 0.89$ ,  $SD = 0.90$ ). Higher scores indicate that the respondent's comment contains more than a singular hateful element. For example, a participant was assigned the hateful condition and wrote the following comment under the "Banana Eating Video": "In our country, refugees live in better conditions than us and this cannot be ignored." Since this comment includes misinformation biased against the refugee, 1 point was assigned to it. The same participant typed, "A wall is not a solution. An electric fence system should be built." under the "Wall on Iran Border" news. Since refugees are seen as a group not worthy of humane treatment, it was identified as dehumanization and given 1 point. In total, this participant's score for making hateful comments was calculated as 2. Each participant's comments were scored in this way to obtain the dependent variable.

### **3.5.3. Independent variable**

The independent variables in my analysis are the experimental conditions: counterspeech and hate speech. To measure the frequency of the independent variables, a categorical scale was developed. "1" indicates participants were assigned the hate speech condition and "2" for the counterspeech condition. At the beginning of the study, it was planned to create three conditions: two treatment groups and a control group that would present the news with neutral comments. After analyzing more than 500 comments on Twitter, Ekşi Sözlük, Facebook, and YouTube under the news, I encountered only one neutral comment pointing out that the headline was wrong.

Of the 181 participants, 94 encountered only hate speech conditions, while 87 were under the counterspeech condition using the Qualtrics tool without any additional intervention. While the number of women exposed to the hate speech condition was 54, the number of women exposed to the counterspeech condition was 37. While under the hate speech condition, 40 persons identified themselves as men; this number was 50 for the counterspeech condition.

## 4. RESULT

### 4.1 Characteristics of the sample

Fifty-one percent of the total respondents ( $N=181$ ) were subjected to the counterspeech condition ( $N_{\text{Counterspeech}}=87$ ). While 42.5% were women under the counterspeech condition, this percentage increased to 57.4% for the hate speech condition. The distribution of ethnic identity between the conditions is well balanced. While 16 individuals under the counterspeech condition described themselves as Kurdish, 62 responders were Turkish. This result was similar under the hate speech condition ( $N_{\text{Kurds}}=18$ ;  $N_{\text{Turks}}=65$ ).

### 4.2 Effect of Counterspeech Condition

To answer the research question, an independent t-test was conducted using SPSS Statistic version 26.0. The result of the test shows that the 94 participants who were exposed to the hate speech condition ( $M=1.04$ ,  $SD=0.92$ ) compared to the 87 participants in the counterspeech condition ( $M=0.73$ ,  $SD=0.86$ ) wrote significantly more hate speech,  $t(179) = 2.2$ ,  $p = .02$ . This result was found even though there was a statistically significant relationship  $t(178) = 1$ ,  $p = .00$  between writing hate speech about refugees and participants' preexisting attitudes toward refugees ( $M=3.32$ ,  $SD=1.06$ ).

In order to further check the robustness of my result, I analyzed the impact of the counterspeech condition after controlling participants' preexisting attitudes toward refugees. A one-way analysis of covariance (ANCOVA) shows that after controlling participants' preexisting attitudes toward refugees, there is a statistically significant difference between making comments about refugees and the content of the exposed comment (the conditions) ( $M_{\text{Condition}}=3.36$ ,  $SD_{\text{condition}}=9.42$ ),  $F(1, 178) = 227.20$ ,  $p < .05$ ,  $\eta^2 = 0.56$ . In other words, the participants exposed to empathy-based counterspeech comments generated less hateful comments than those exposed to hateful comments:

( $M_{counterspeech}=0.73$ ,  $SD_{counterspeech}=0.86$ ), ( $M_{hatespeech}=1.04$ ,  $SD_{hatespeech}=0.92$ ),  $F(1, 178) = 9.42$ ,  $p < .05$ ,  $\eta^2 = .05$ .

### 4.3 The Effects of Political Party Voting on Writing Hate Speech

To analyze the effect of political party voting on writing hate speech about refugees, the Kruskal-Wallis Test showed a statistically significant difference between writing hate speech comments and political party voting.  $H(6)=49.87$ ,  $p=.00$ . On a party basis, the Games-Howell Test showed a significant difference between voters of HDP  $p=.00$  and every other party, except for other voters  $p=.91$ .

To demonstrate the robustness of the result, I applied the Non-Parametric ANCOVA (Quade's) test after controlling participants' preexisting attitudes. The test Quade's also shows a statistically significant difference between writing hate speech comments and political party voting.  $F(6, 174)= 2.904$ ,  $p=0.01$ . Compared to the Games-Howell Test, the only statistically significant difference was found between participants who voted for AKP and HDP,  $p=0.00$ . In other words, HPD voters wrote less hate speech comments than AKP supporters.

### 4.4 The Effect of Ethnic Identification on Writing Hate Speech

In order to understand the effect of ethnic identity on writing hate speech, the Kruskal-Wallis Test showed a statistically significant difference between writing hate speech comments and ethnic identity  $H(3)=38.95$ ,  $p=0.00$ . Moreover, according to the Games-Howell Test, the mean value of writing hate speech comments is significantly different between Kurd and Turk ( $p=0.00$ ) as well as between Turk and others ( $p=0.00$ ). In other words, Kurd wrote less hate speech comments ( $M=.17$ ,  $SD=.45$ ) compared to Turk ( $M=1.16$ ,  $SD=.88$ ).

After controlling participants' preexisting attitudes toward refugees, the result was not robust. The Non-Parametric ANCOVA (Quade's) test showed no statistically significant difference between writing hate speech and ethnic identity.  $F(3, 177)= 2.62$ ,  $p=.05$  after controlling participants' preexisting attitudes toward refugees. However, there was a significant difference between Kurd and Turk,  $p=.03$ .

#### **4.5. Analyzing Open-Ended Comments**

According to content analysis of responses to the open-ended survey questions on posting comments related to shown news, 128 of 364 comments included hostile messaging about refugees. Twenty-one percent of the hateful comments were coded under the categories of dehumanization, mostly characterizing refugees as people who are undesirable in Turkey. 18% of the hateful comments were coded under the insulting categories in which the most encountered words were "uneducated," "unqualified," and "stateless." Sixteen percent of the hateful comments were categorized under misinformation. The most common misinformation was assumed that refugees "live in comfort with the taxes of locals " and different theories about the presence of refugees in Turkey. Finally, with the highest percentage, 42% of hateful comments were coded under the category of othering. The distinction between "us" and "them" was one of the most frequently encountered discourses of othering. In addition, the perception of refugees as a threat to the way of life of locals and demands for restricting their access to basic rights were among the comments frequently encountered under the category of othering.

#### **4.6 Analyzing Preexisting Perceptions of Participants Toward Refugees**

Based on the survey analysis, it was revealed that people living in Turkey have negative perceptions of refugees since among the total answers to the five questions containing negative perceptions towards refugees, approximately 50% of the responses were marked as strongly agree or agree. For example, the survey found that 47,5 of the participants do not want more refugees to come to Turkey and demand the closure of the borders.

It was found that while only 4,4% of participants think that refugees increase Turkey's cultural diversity, 28,7 % see refugees as a reason for the increasing rate of criminal offenses. Furthermore, the respondents said that (%47) refugees should not have equal rights as Turkish citizens. Below you may see the **Table 4.1** for the detailed responses of the participants.

**Table 4.1***Percentage of Preexisting Perceptions of Participants Toward Refugees*

| Questions  | Strongly disagree | Disagree | Both agree and disagree | Agree | Strongly agree |
|--|-------------------|----------|-------------------------|-------|----------------|
| Turkey hosts the most significant number of refugees in the world. Therefore, it should not take in any more refugees, and the borders should be closed. | 13,3              | 19,9     | 19,3                    | 11,6  | 35,9           |
| Most refugees who want to come to Turkey are not fleeing war. They come here for economic reasons.   | 12,2              | 22,7     | 22,7                    | 24,9  | 17,7           |
| Refugees quickly adapt to Turkey's culture, which increases our cultural diversity.  | 37,6              | 20,4     | 22,1                    | 15,5  | 4,4            |
| The judicial crime rate in Turkey has increased due to the growth of the number of refugees.   | 9,4               | 18,2     | 21,5                    | 21,1  | 28,7           |
| The opening of new workplaces by refugees helps the Turkish economy to grow.   | 29,3              | 27,1     | 24,9                    | 16    | 2,8            |
| One of the biggest reasons for the recent housing problems is the refugees coming to our country.  | 15,5              | 23,8     | 16                      | 23,8  | 21             |
| Refugees are one of the reasons why the unemployment rate in Turkey is so high.  | 20,4              | 27,1     | 13,3                    | 17,1  | 22,1           |
| Refugees living in Turkey work and pay taxes. Therefore, they should have the same rights as Turkish citizens.   | 29,3              | 17,7     | 16                      | 27,6  | 9,4            |

It also highlighted that Kurds ( $M_{Kurds}=2,54$ ,  $SD_{Kurds}=1,22$ ) have less negative perceptions toward refugees than Turks ( $M_{Turks}=3,63$ ,  $SD_{Turks}=1,28$ ). Since women are also one of the vulnerable groups in Turkey, women were expected to have less negative perceptions toward refugees than men. However, the result was the opposite of this expectation. The survey showed that women ( $M_{Women}=3,34$ ,  $SD_{Women}=1,29$ ) have more negative attitudes toward refugees than men ( $M_{Men}=3,31$ ,  $SD_{Men}=1,38$ ).

## 5. DISCUSSION

While there is growing literature on the consequences of online hate speech and its adverse impacts on affected communities, there is dearth of studies examining methods to mitigate it and their effectiveness (Hangartner et al. 2021). Yet, one of the methods that has not been studied substantially is counterspeech, which is increasingly used to combat online hate speech (Buerger 2021a). Systematic examination of the efficacy of counterspeech strategies is acutely lacking. This study is aimed to investigate how exposure to solely empathy-based counterspeech comments about refugees makes a difference in social media users' comments on Twitter conversation in Turkey. To that end, an online survey experiment with 181 adults from Turkey was conducted. It was found that participants were less likely to make hostile comments when presented with an environment in which they were only exposed to empathy-based counterspeech comments. This outcome corresponds to similar findings of Hangartner (2021) which found that only empathy-based counterspeech was effective in changing the behavior of hate speech generators among the various strategies such as warning of consequences, and humor.

Refugees are at the center of online hate speech in Turkey; the content and language used in online environments further strengthen hate speech directed at them (Alikılıç et al. 2021, 501). In total, 63.6 % of the comments about refugees on Twitter have negative connotations (510). Similar attitudes to these in Turkey are also being observed worldwide. Šori and Vehovar (2022) found that refugees are one of the vulnerable groups that are most exposed to online hate speech. Counterspeech, an effective method to combat hate speech against different minority groups, can also be used to combat online hate speech against refugees. For example, research conducted by Garland et al. (2020) shows that the organized counterspeech group RI has succeeded in changing hostile perceptions of RQ group members toward minorities in Germany. This study obtained a similar result to the existing, albeit limited, literature on this issue. It was found that although there is a significant correlation between preexisting attitudes towards refugees and the generation of hostile social media comments, participants who are exposed to only counterspeech comments are less likely to share hateful comments



than those who are exposed to hate speech comments. In other words, even if individuals have a negative perception of refugees, they create less hate speech when they encounter comments containing counterspeech.

Other than ethnic identity and political party voting, there is no significant relationship between demographic characteristics and writing hate speech. The finding of the relationship between political party voting and writing hate speech is supported by existing studies. Brooks, Manza and Cohen (2016) demonstrated that individuals' party identification affected their perception of refugees in the USA. Moreover, Bernatzky, Costello and Hawdon (2021) found that individuals who support Donald Trump are more likely to generate hate online. Other studies suggested that social polarization, one of the most crucial sources of hate speech (Hrant Dink Foundation 2019, 85), is also correlated with party identification (Pérez-Escolar and Noguera-Vivo 2022, 206). Erdoğan and Uyan-Semerci's (2022) study states that polarized groups come together when it comes to refugees and perceive refugees as a common "enemy" and "other". However, contrary to Erdoğan and Uyan-Semerci's study, the data reported in this study indicates that significant differences in hate speech generation exist among the participants who voted for AKP and HDP, which are the most polarized groups in Turkey (TurkuazLab 2020, 2). Turkey's political polarization is one of the main factors affecting negative attitudes toward refugees, mainly Syrian (Morgül et al. 2021, 14). Studies have shown that counterspeech, used to combat online hate speech, also contributes to depolarization. For example, a study conducted by a research team claims that particularly organized counterspeech could balance and decrease polarization (Garland et al. 2020, 110). Since the study reported provides evidence that counterspeech is an effective method to combat online hate speech and considering previous studies related to counterspeech and polarization, it also might be argued that counterspeech plays an essential role in reducing polarization.

According to research carried out by Turkish Economic Social and Political Research Foundation (Türkiye Sosyal Ekonomik ve Siyasal Araştırmalar Vakfı -TüSES), Kurds are more moderate towards refugees than the Turkish majority (Morgül et al. 2021, 15). Their research has also suggested that ethnic and religious minorities other than Kurds

and Alevi generally have negative feelings about refugees, mainly Syrians (16). Similar to TüSES's arguments, this study also found that although statistically significant results do not exist between writing hate speech about refugees and ethnicity after controlling participants' preexisting attitudes, there was a correlation in Kurdish and Turkish participants; Kurdish participants generate fewer hateful comments than Turkish ones. It is unsurprising that Kurds, being subjected to various forms of social pressure (Morgül et al. 2021, 71) and being an important target of hate speech (Hrant Dink Foundation 2010, 5), produce fewer hate speech comments against refugees compared to Turks.

A limitation of the study is related to the sample size. The external validity of the experiments testing the effect of counterspeech can be improved with bigger samples. This would allow the inclusion of more control variables to test the alternative explanations and rule them out wherever necessary.

Regarding further research, studies dealing with factors that determine the effectiveness of counterspeech on different social media platforms such as Facebook, TikTok, Instagram should be encouraged. Specifically, how the tone of counterspeech messaging, (e.g., generating civil or uncivil messaging and characteristics of counter-speakers such as race, gender, age, ethnicity, etc.,) affects the impact of counterspeech should be examined. Moreover, the effect of the counter speakers' race, gender, and ethnic identity in generating more counterspeech can also be studied. In other words, counterspeech is generated more when it comes from whomever. Since this study is the first survey experiment in Turkey on counterspeech, I suggest more empirical studies in which researchers could test whether counterspeech might decrease online hate speech through long-term observations. Furthermore, these studies might eventually fill the gaps in the research literature on the ecosystem of hate speech in Turkey by understanding the cause and effect of hate speech generation of social media users, as well as its relationships with different variables, such as social and political identities, sociodemographic factors. As counterspeech is not only used to combat online hate speech, but also various social problems, (e.g., polarization) researchers should be encouraged to evaluate how counterspeech may be an effective strategy to counter other negative social attitudes in Turkey besides hate speech. Finally, considering that

refugees are not the only vulnerable groups that face online hate speech, similar studies may be conducted taking into consideration a variety of minorities such as LGBTI+, Kurds, Alevi, etc.



## BIBLIOGRAPHY

- AliKılıç, Özlem, Ebru Gökaliler, and İnanç AliKılıç. 2021. “Nefret Söylemi Üzerinden Ötekileştirme: Twitter’da Mültecilere Yönelik Nefret Tipolojisi Analizi.” *Akdeniz İletişim* 0, no.36: 501-520. <https://doi.org/10.31123/akil.989074>
- Allan, James. 2013. “Hate Speech Law and Disagreement.” *University of Minnesota Law School* 29, no.1 (Summer): 59-79.
- Álvarez-Benjumea, Amalia, and Fabian Winter. 2018. “Normative Change and Culture of Hate: An Experiment in Online Environments.” *European Sociological Review* 34, no.3: 223–237. <https://doi.org/10.1093/esr/jcy005>.
- Akgül, Mahmut. 2020. “Çevrimiçi Ortamlarda Nefret Söylemi: Ekşi Sözlük’te 65 Yaş Üstü Sokağa Çıkma Yasağı Tartışmaları.” *İletişim Kuram ve Araştırma Dergisi* 2020, no. 51: 57–78.
- Aslan, Alev. 2018. “Online Hate Discourse: A Study on Hatred Speech Directed Against Syrian Refugees on YouTube.” *Journal of Media Critiques* 3, no.12: 227–56. <https://doi.org/10.17349/jmc117413>.
- Bakalis Chara. 2015. *Cyberhate: An Issue of Continued Concern for the Council of Europe’s Anti-Racism Commission*. Ref. 207215GBR. Strasbourg Cedex, France: Council of European. <https://edoc.coe.int/en/cybercrime/6883-cyberhate-an-issue-of-continued-concern-for-the-council-of-europe-s-anti-racism-commission.html>.
- Baker, Edwin C. 2008. “Hate Speech.” *Faculty Scholarship at Penn Law*, no.198: 1-23. [https://scholarship.law.upenn.edu/faculty\\_scholarship/198](https://scholarship.law.upenn.edu/faculty_scholarship/198).
- Bartlett, Jamie, and Alex Krasodonski-Jones. 2015. *Counter-Speech: Examining Content That Challenges Extremism Online*. London, The United Kingdom: Demos. <https://demos.co.uk/project/counter-speech/>.
- Benesch, Susan, Catherine Buerger, and Sean Manion. 2018. *Dangerous Speech: A Practical Guide Anthropology of Human Rights View Project*. Washington, The United State of America. The <https://dangerousspeech.org/guide/>.
- Benesch, Susan. 2020. “Countering Dangerous Speech: New Ideas for Genocide Prevention.” Working Paper, United States Holocaust Memorial Museum. <https://www.ushmm.org/m/pdfs/20140212-benesch-countering-dangerous-speech.pdf>.
- Bernatzky, Colin, Matthew Costello, and James Hawdon. 2021. “Who Produces Online Hate?: An Examination of the Effects of Self-Control, Social Structure, & Social Learning.” *American Journal of Criminal Justice* 47, no.3: 421–40. <https://doi.org/10.1007/s12103-020-09597-3>.
- Bianet*. 2021. “Facing deportation over news on ‘banana-eating videos’, journalist

Shamaa released.” November 9, 2021. <https://m.bianet.org/english/migration/253054-facing-deportation-over-news-on-banana-eating-videos-journalist-shamaa-released>.

Binark, Mutlu, and Tuğrul Çomu. 2012. “Sosyal Medyanın Nefret Söylemi İçin Kullanılması İfade Özgürlüğü Değildir!” Accessed on September 15, 2022. <https://yenimedya.wordpress.com/2012/01/20/sosyal-medyanin-nefret-soylemi-icin-kullanilmasi-ifade-ozgurlugu-degildir/>.

Brown, Alexander. 2017. “What Is Hate Speech? Part 1: The Myth of Hate.” *Law and Philosophy* 36, no.4: 419–468. <https://doi.org/10.1007/s10982-017-9297-1>.

Brown, Alexander. 2018. “What Is so Special About Online (as Compared to Offline) Hate Speech?” *Ethnicities* 18, no.3: 297–326. <https://doi.org/10.1177/1468796817709846>.

Buerger, Cathy. 2021a. “Counterspeech: A Literature Review.” Accessed on September 15, 2022. <http://dx.doi.org/10.2139/ssrn.4066882>.

Buerger, Catherine. 2021b. “#iamhere: Collective Counterspeech and the Quest to Improve Online Discourse.” *Social Media + Society* 7, no. 4: 205630512110638. <https://doi.org/10.1177/20563051211063843>.

Cepollaro, Bianca, Lepoutre, Maxime, and Robert Mark Simpson. 2022. “Counterspeech.” *Philosophy Compass* 2022, no.e12890: 1-11. <https://doi.org/10.1111/phc3.12890>.

Chaudhary, Mudit, Chandni Saxena, and Helen Meng. 2021. “Countering Online Hate Speech: An NLP Perspective.” <https://doi.org/10.48550/ARXIV.2109.02941>, <https://arxiv.org/abs/2109.029417>.

Cheng, Justin, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. “Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions.” In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1217–30. Portland Oregon, USA: ACM. <https://doi.org/10.1145/2998181.2998213>.

Cinelli, Matteo, Andraž Pelicon, Igor Mozetič, Walter Quattrociocchi, Petra Kralj Novak, and Fabiana Zollo. 2021. “Dynamics of Online Hate and Misinformation.” *Scientific Reports* 11, no.1: 1–13. <https://doi.org/10.1038/s41598-021-01487-w>.

Cohen-Almagor, Raphael. 2015. *Confronting the Internet’s Dark Side: Moral and Social Responsibility on the Free Highway*. Cambridge, UK: Cambridge University Press.

Council Of Europe Committee of Ministers. 1997. “Recommendation No. R (97) 20.” In *Freedom of Expression and the Media: Standard Setting by Council of European*, edited by Susanne Nikoltchev, 106–8. Strasbourg: European Audiovisual Observatory

Dangerous Speech Project. n.d. “Countering Dangerous Speech Around the World.” Accessed September 15, 2022. <https://dangerousspeech.org/countries/>.

De Gibert, Ona, Naiara Perez, Aitor García-Pablos, and Montse Cuadros. 2018. "Hate Speech Dataset from a White Supremacy Forum." In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, 11–20, Brussels, Belgium. Association for Computational Linguistics.

Diaz Gustavo, Grady Christopher, and Kuklinski James H. 2020. "Survey Experiments and the Quest for Valid Interpretation." In *The SAGE Handbook of Research Methods in Political Science and International Relations*, edited by Luigi Curini and Franzese R, 1036–1052. California: SAGE Publications.

Dondurucu, Zeynep Benan. 2018a. "Yeni Medyada Cinsel Kimlik Temelli Nefret Söylemi: İnci Sözlük Örneği." *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi* 6, no.2: 1376-1405. <https://doi.org/10.19145/e-gifder.435744>.

Dondurucu, Zeynep Benan. 2018b. "Eşcinsellik Temelli Nefret Söylemi İçeren İletilerin Twitter'da İncelenmesi." *Erciyes İletişim Dergisi* 5, no.4: 513–34. <https://doi.org/10.17680/erciyesiletisim.420520>.

English, Morgan. 2021. "Cancel Culture: An Examination of Cancel Culture Acts as a Form of Counterspeech to Regulate Hate Speech Online." Master diss., University of North Carolina.

Erbaysal-Filibeli, Tirşe and Can Ertuna. 2021. "Sarcasm Beyond Hate Speech: Facebook Comments on Syrian Refugees in Turkey." *International Journal of Communication* 15: 2236–59.

Erdoğan, Emre, and Pınar Uyan-Semerci. 2022. *Kutuplaşmayı Nasıl Aşarız?*. İstanbul: turkuazlab.

Erdoğan-Öztürk, Yasemin, and Hale Işık-Güler. 2020. "Discourses of Exclusion on Twitter in the Turkish Context: #ülkemdesuriyeliistemiyorum (#idontwantsyriansinmycountry)." *Discourse, Context and Media* 36 (August): 1–10. <https://doi.org/10.1016/j.dcm.2020.100400>.

Friess, Dennis, Marc Ziegele, and Dominique Heinbach. 2020. "Collective Civic Moderation for Deliberation? Exploring the Links between Citizens' Organized Engagement in Comment Sections and the Deliberative Quality of Online Discussions." *Routledge* 38, no.5: 624–46. <https://doi.org/10.1080/10584609.2020.1830322>.

Gagliardone, Iginio. 2014. "Mapping and Analysing Hate Speech Online." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2601792>

Garland, Joshua, Keyan Ghazi-Zahedi, Jean-Gabriel Young, Laurent Hébert-Dufresne, and Mirta Galesic. 2020. "Countering Hate on Social Media: Large Scale Classification of Hate and Counter Speech." In *Proceedings of the fourth workshop on online abuse and harms. Association for Computational Linguistics*, 102–112. <https://www.aclweb.org/anthology/2020.alw-1.13>.

Gelashvili, Teona. 2018. "Hate Speech on Social Media: Implications of private regulation and governance gaps." Master's thesis. Lund University.

<https://lup.lub.lu.se/luur/download?func=downloadFile&recordOid=8952399&fileOid=8952403>.

Grabe, Maria, Özen Bas, Louis Pagano, and Lelia Samson. 2012. "The Architecture of Female Competition: Derogation of a Sexualized Female News Anchor." *Journal of Evolutionary Psychology* 10, no.3: 107–33. <https://doi.org/10.1556/JEP.10.2012.3.2>.

Han, Soo Hye, and Le Ann M. Brazeal. 2015. "Playing Nice: Modeling Civility in Online Political Discussions." *Communication Research Reports* 32, no.1: 20–28. <https://doi.org/10.1080/08824096.2014.989971>.

Hangartner, Dominik, Gloria Gennaro, Sary Alasiri, Nicholas Bahrach, Alexandra Bornhoft, Joseph Boucher, Buket Buse Demirci, et al. 2021. "Empathy-Based Counterspeech Can Reduce Racist Hate Speech in a Social Media Field Experiment." *Proceedings of the National Academy of Sciences of the United States of America* 118 no.50: 1–3. <https://doi.org/10.1073/pnas.2116310118>.

Hawdon, James, Atte Oksanen, and Pekka Räsänen. 2017. "Exposure to Online Hate in Four Nations: A Cross-National Consideration." *Deviant Behavior* 38, no.3: 254–66. <https://doi.org/10.1080/01639625.2016.1196985>.

Henson, Billy, Bonnie S. Fisher, and Bradford W. Reynolds. 2020. "There Is Virtually No Excuse: The Frequency and Predictors of College Students' Bystander Intervention Behaviors Directed at Online Victimization." *Violence Against Women* 26, no.5: 505–27. <https://doi.org/10.1177/1077801219835050>.

Hrant Dink Foundation. 2019. *Medyada Nefret Söylemi ve Ayrımcı Söylem 2019 Raporu*. Report No: 9786058071254. <https://hrantdink.org/tr/asulis/yayinlar/72-medyada-nefret-soylemi-raporlari/2665-medyada-nefret-soylemi-ve-ayrimci-soylem-2019-raporu>.

Hrant Dink Foundation. 2014. *Medyada Nefret Söylemi ve Ayrımcı Dil Eylül - Aralık 2014 Raporu*. <https://hrantdink.org/tr/asulis/yayinlar/72-medyada-nefret-soylemi-raporlari/394-medyada-nefret-soylemi-ve-ayrimci-dil-eylul-aralik-2014>

Hürriyet Daily News. 2021. *Turkey Extends Security Wall along Iran Border: Interior Minister*. Accessed September 15, 2022. <https://www.hurriyetaidailynews.com/turkey-extends-security-wall-along-iran-border-interior-minister-167897>.

İnsan Hakları Derneği. 2020. *Türkiye'de Nefret Suçları ve Son Dönemde Yaşanan İrkçi Saldırıları Özel Raporu*. <https://www.ihd.org.tr/turkiyede-nefret-suclari-ve-son-donemde-yasanan-irkci-saldirlar-ozel-raporu/>

Jakubowicz, Andrew, Kevin Dunn, Gail Mason, Yin Paradies, Ana-Maria Bliuc, Nasya Bahfen, Andre Oboler, Rosalie Atie, and Karen Connelly. 2017. *Cyber Racism and Community Resilience Strategies for Combating Online Race Hate*. Cham: Springer International Publishing AG.

Kaos GL. 2020. *Medya İzleme 2020 Raporu*. Accessed September 15, 2022 <https://kaosgldernegi.org/images/library/medya-i-zleme-2020-web.pdf>

Kaos GL. 2021. "Dört Maddede Nefret Söylemi Nedir?" 17.02.2020. <https://kaosgl.org/haber/dort-maddede-nefret-soylemi-nedir>

Keller, Nadide, and Tina Askanus. 2020. "Combatting hate and trolling with love and reason?: a qualitative analysis of the discursive antagonisms between organised hate speech and counterspeech online." *Studies in Communication and Media* 9, no.4: p. 540-572. <https://doi.org/10.5771/2192-4007-2020-4-540>.

Kennedy, Brendan, Kristopher Coombs, G J Portillo-Wightman, Aida Mostafazadeh Davani, Drew Kogon, Kris Coombs, Joe Hoover, et al. 2018. "A Typology and Coding Manual for the Study of Hate-Based Rhetoric." *PsyArXiv*. 1–20.

Kim, Jae Yeon, Jaeung Sim, and Daegon Cho. 2022. "Identity and Status: When Counterspeech Increases Hate Speech Reporting and Why." *Information Systems Frontiers* 10229, no.2: 3–12. <https://doi.org/10.1007/s10796-021-10229-2>.

Kınay, Kumsal, and Esra Gül Atalay. 2021. "Sosyal Medyada Trol Hesaplar ve Algı Yönetimi: COVID-19 Sürecinde Twitter'da Aşı Karşıtlığı." *Medya ve Kültüre Çalışmalar Dergisi* 3, no.2: 56–63. <https://doi.org/10.29228/mekcad.11>.

Klein, Adam. 2012. "Slipping Racism into the Mainstream: A Theory of Information Laundering." *Communication Theory* 22, no.4: 427–48. <https://doi.org/10.1111/j.1468-2885.2012.01415.x>.

Konda Araştırma ve Danışmanlık. 2022. "Türkiye 100 Kişi Olsaydı." Accessed September 15, 2022. <https://interaktif.konda.com.tr/turkiye-100-kisi-olsaydi>.

Latour, Agata de, Nina Perger, Ron Salaj Claudio Tocchi and Paloma Viejo Otero. 2017. *Taking Action Against Hate Speech Through Counter and Alternative Narratives*. Report No. 9789287184450. <https://rm.coe.int/wecan-eng-final-23052017-web/168071ba08>.

MacAvaney, Sean, Hao Ren Yao, Eugene Yang, Katina Russell, Nazli Goharian, and Ophir Frieder. 2019. "Hate Speech Detection: Challenges and Solutions." *PLoS ONE* 14, no.8: 1-16. <https://doi.org/10.1371/journal.pone.0221152>.

McDoom, Omar Shahabudin. 2012. "The Psychology of Threat in Intergroup Conict: Emotions, Rationality, and Opportunity in the Rwandan Genocide." *International Security* 37, no. 2: 119–55. [https://doi.org/10.1162/ISEC\\_a\\_00100](https://doi.org/10.1162/ISEC_a_00100).

Miškolci, Jozef, Lucia Kováčová, and Edita Rigová. 2018. "Countering Hate Speech on Facebook: The Case of the Roma Minority in Slovakia." *Social Science Computer Review* 38, no. 2: 128–46. <https://doi.org/10.1177/0894439318791786>.

Molina, Rocío Galarza, and Freddie J. Jennings. 2018. "The Role of Civility and Metacommunication in Facebook Discussions." *Communication Studies* 69, no. 1: 42–66. <https://doi.org/10.1080/10510974.2017.1397038>.

Müller, Karsten and Carlo Schwarz. 2021. "Fanning the Flames of Hate: Social Media and Hate Crime." *Journal of the European Economic Association* 19, no.4: 2131–67.



<https://doi.org/10.1093/jeea/jvaa045>.

NTV. 2014. “Kahramanmaraş’ta Suriyeli Gerginliği.” July 13, 2014. <https://www.ntv.com.tr/galeri/turkiye/kahramanmarasta-suriyeli-gerginligi,BZ47FCFQHEGz4WT1PwoO1Q>.

Obermaier, Magdalena, Desirée Schmuck, and Muniba Saleem. 2021. “I’ll Be There for You? Effects of Islamophobia Online Hate Speech and Counter Speech on Muslim in-Group Bystanders’ Intention to Intervene.” *New Media and Society* 00, no.0: 1-20. <https://doi.org/10.1177/14614448211017527>.

Önder, Mehmet Seyman. 2020. “Mültecilerin Etiketlenmesi Ölçeği: Geçerlik ve Güvenirlik Çalışması.” *Journal of Turkish Studies* 15, no. 5: 2545–61. <https://doi.org/10.7827/turkishstudies.43848>.

Özatalay, Cem, and Seçil Doğuş. 2018. *Türkiye’de Ayrımcılık Algısı*. İstanbul: Eşit Haklar için İzleme Derneği.

Pak, Halil. 2020. “Socioeconomic Conflict between Host Community and Syrian Refugees in Urban Turkey: The Mediating Role of Political Trust.” *Studies in Psychology* 40, no. 2: 579–97. <https://doi.org/10.26650/sp2019-0094>.

Reboot. n.d. “The Social Platforms with the Biggest Increase in Removed Content | 2019-2020.” Accessed September 15, 2022. <https://www.rebootonline.com/digital-pr/assets/social-media-platforms-biggest-increase-removed-content/>.

Resmi Gazete. 2020. “İnternet Ortamında Yapılan Yayınların Düzenlenmesi ve Bu Yayınlar Yoluyla İşlenen Suçlarla Mücadele Edilmesi Hakkında Kanunda Değişiklik Yapılmasına Dair Kanun.” Accessed September 15, 2022. <https://www.resmigazete.gov.tr/eskiler/2020/07/20200731-1.htm>

Pérez-Escobar, Marta, and José Manuel Noguera-Vivo. 2021. *Hate Speech and Polarization in Participatory Society*. London: Routledge. <https://doi.org/10.4324/9781003109891>.

Rudnicki, Konrad, and Stefan Steiger. 2020. *Online Hate Speech-Introduction into Motivational Causes, Effects and Regulatory Contexts*. Detect Then ACT. <https://www.media-diversity.org/resources/online-hate-speech-introduction-into-motivational-causes-effects-and-regulatory-context/>.

Polat, Burak, Necmi Emel DiLmen, and Cem Sefa Sütçü. 2017. “Türkiye’de Twitter Kullanıcılarının Retweet Pratikleri Üzerine Kullanımlar ve Doyumlar Paradigması ile Bir Karma Araştırma.” *Electronic Journal of New Media* 5, no. 2: 112–35. [https://doi.org/10.17932/IAU.EJNM.25480200.2021/ejnm\\_v5i2002](https://doi.org/10.17932/IAU.EJNM.25480200.2021/ejnm_v5i2002).

Saha, Punyajoy, Binny Mathew, Pawan Goyal, and Animesh Mukherjee. 2018. “Hateminers: Detecting Hate Speech Against Women.” *ArXiv Preprint arXiv.1812.06700*. 1–5. <https://doi.org/10.48550/arXiv.1812.06700>.

Schieb, Carla, and Mike Preuss. 2016. “Governing Hate Speech by Means of

Counterspeech on Facebook.” In *66th ICA Annual Conference* 1-23.

Schnabel, Landon. 2021. “Survey Experiments.” In *The Routledge Handbook of Research Methods in the Study of Religion*, edited by Steven Engler, Michael Stausberg, 3-21. Valencia: Association for Computational Linguistics.

Sellars, Andrew F. 2016. “Defining Hate Speech.” *Berkman Klein Center Research Publication* 2016, no.20: 5–31.

Šori, Iztok, and Vasja Vehovar. 2022. “Reported User-Generated Online Hate Speech: The ‘Ecosystem’, Frames, and Ideologies.” *Social Sciences* 11, no. 8: 375. <https://doi.org/10.3390/socsci11080375>.

Southern Poverty Law Center. 2022. *The Year in Hate & Extremism 2021*. Accessed September 15, 2022. <https://www.splcenter.org/year-hate-extremism-2021>.

Strand, C., J. Svensson, R. Blomeyer, and M. Sanz. 2021. *Disinformation Campaigns about LGBTI+ People in the EU and Foreign Influence*. Report No: 978-92-846-8347-5. Brussel, Belgium: European Union.

Stroud, Scott R., and William Cox. 2018. “The Varieties of Feminist Counterspeech in the Misogynistic Online World.” In *Mediating Misogyny Gender, Technology, and Harassment* edited by Jacqueline Ryan Vickery and Tracy Everbach, 293-310. Basingstoke, UK: Palgrave Macmillan.

Taşdelen, Birgül. 2020. “Twitter’da Suriyeli Mültecilere Karşı Çevrimiçi Nefret Söylemi:#suriyelileriistemiyoruz.” *Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Elektronik Dergisi* 11, no.2: 561–75.

Teyit. 2019a. *Medya Kullanımı ve Haber Tüketimi: Güven, Doğrulama, Siyasi Kutuplaşmalar*. Accessed September 15, 2022. <https://cdn.teyit.org/wp-content/uploads/2019/01/medya-kullanimi-ve-haber-tuketimi-teyit-ocak-2019.pdf>.

Teyit. 2019b. “Yavuz Sultan Selim Üniversitesi’nde Yazılan Doktora Tezi İçin Suriyeli Bir Kişinin Verdiği Demeçle İlgili İddia.” November 5, 2019. <https://teyit.org/yavuz-sultan-selim-universitesinde-yazilan-doktora-tezi-icin-suriyeli-bir-kisinin-verdigi-demecele-ilgili-iddia>.

Türkiye İstatistik Kurumu. 2022. “İstatistiklerle Yaşlılar 2021”. Lasted modified March 18, 2022. <https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Yasli-lar-2021-45636>.

The UN Refugee Agency. n.d. " What is a refugee?." Accessed December 26, 2022. <https://www.unhcr.org/what-is-a-refugee.html>.

Twitter. 2021. *Twitter Türkiye Şeffaflık Raporu Haziran 2021*. Accessed September 15, 2022. <https://help.twitter.com/de/rules-and-policies/enforcement-philosophy>.

Twitter. n.d. “Hateful Conduct Policy.” Accessed September 15, 2022. <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>.

United Nations. 2019. “United Nations Strategy and Plan of Action on Hate Speech.” United Nations Report. June 18, 2019. <https://www.un.org/en/hate-speech/un-strategy-and-plan-of-action-on-hate-speech>.

United Nations Educational, Scientific and Cultural Organization. 2021. *Addressing Hate Speech on Social Media: Contemporary Challenges*. Report No: CI/FEJ/2021/DP/01. <https://unesdoc.unesco.org/ark:/48223/pf0000379177>.

United Nations High Commissioner for Refugees. n.d. “Figures at a Glance.” Accessed September 15, 2022. <https://www.unhcr.org/figures-at-a-glance.html>.

Waltman, Michael S. and A. Ashely Mattheis. 2017. “Understanding hate speech.” In *Oxford Research Encyclopedia of Communication*. <https://doi.org/10.1093/acrefore/9780190228613.013.422>.

Wachs, Sebastian, and Michelle F. Wright. 2018. “Associations between Bystanders and Perpetrators of Online Hate: The Moderating Role of Toxic Online Disinhibition.” *International Journal of Environmental Research and Public Health* 15, no.9: 1–9. <https://doi.org/10.3390/ijerph15092030>.

Weber, Anne. 2009. *Manual on Hate Speech*. France: Council of Europe.

Wright, Lucas, Derek Ruths, Kelly P Dillon, Haji Mohammad Saleem, and Susan Benesch. 2017. “Vectors for Counterspeech on Twitter.” In *2017 Proceedings of the First Workshop on Abusive Language Online* 57–62. <https://doi.org/10.18653/v1/w17-3009>.

Yıldız, Engincan. 2018. “Twitter’da ve Çevrimiçi Bir Gazetede Yer Alan Nefret Söylemlerinin Karşılaştırılması: Suriyeli Mülteciler Örneği.” *OPUS Uluslararası Toplum Araştırmaları Dergisi* 9, no.16: 78–78. <https://doi.org/10.26466/opus.478176>.

YouTube. n.d. “Hate Speech Policy.” Accessed September 15, 2022. <https://support.google.com/youtube/answer/2801939?hl=en>.

## APPENDIX A

### A.1 News and Comments Showed During the Experimental Stimuli

#### A.1.1 Hate speech condition

|                                     |  |
|-------------------------------------|--|
| Banana Eating Videos – Post Caption | In Izmir, after a citizen said in a street interview, "I can't eat bananas. They buy bananas by the kilograms." Following the "banana eating" videos shared by Syrians on social media, 8 Syrians were detained. |
| Comment - 1                         | I don't want refugees in my country.<br>I hope they are all deported.  |
| Comment - 2                         | Just yesterday you were crossing the border crying.<br>Now you are making fun of us. Ingrates!   |
| Comment - 3                         | Do I have to live in the same country with them...<br>Go to your own country.  |
| Comment - 4                         | We have become refugees in our own country.<br>We are the refugees.  |
| Comment - 5                         | Let them live in luxury with our taxes.<br>We can't even afford bananas.   |

|                                    |  |
|------------------------------------|--|
| Wall on Iran Border – Post Caption | Van Governor Bilmez: "A wall will be built along the entire 295-kilometer Iranian border to prevent illegal crossings."  |
| Comment - 1                        | It's early!!! The other half of Afghanistan has not arrived yet!!! When they arrive, then we will raise the wall...  |
| Comment - 2                        | For God's sake, don't make these walls 2 meters tall. If the footage I saw is real, jumping over it is a piece of cake.  |
| Comment - 3                        | WALLS ARE NOT THE SOLUTION!! Men are jumping on the wing of a moving airplane. We need a permanent solution.   |
| Comment - 4                        | A man walks 1000 km to reach the border...Will the wall stop him... Let them come too, after all we are the refugee guardians of Europe.                                 |
| Comment - 5                        | I'm not saying refugees shouldn't come to Turkey, but they need to be taken from the inner cities to the camps. Our biggest mistake was taking them to the city centers. |

#### A.1.2 Counterspeech Condition

|                                     |  |
|-------------------------------------|--|
| Banana Eating Videos – Post Caption | In Izmir, after a citizen said in a street interview, "I can't eat bananas. They buy bananas by the kilograms." Following the "banana eating" videos shared by Syrians on social media, 8 Syrians were detained. |
|-------------------------------------|--|

|             |   |
|-------------|---|
| Comment - 1 | I'm so sorry to hear that. So what? These people have exercised their right to protest, just like the rest of us. For real justice to come, we must demand justice for all of us.                 |
| Comment - 2 | We cannot blame all refugees for what one group did. One day we may become refugees ourselves, we may have to leave our country. And on that day, the food we eat will also sting someone's eyes. |
| Comment - 3 | Politicians and governments have turned these people into monsters in the eyes of all of us... They are human beings just like us and they deserve to live humanely.                              |
| Comment - 4 | If you are against helping helpless people, then you deserve to be governed like this. First ask not why Syrians can eat bananas but why we can't...  |
| Comment - 5 | 72% of young people in Turkey want to go to Europe. I wonder if Europe will see a banana as too much for our children.  |

|                                    |   |
|------------------------------------|---|
| Wall on Iran Border – Post Caption | Van Governor Bilmez: "A wall will be built along the entire 295-kilometer Iranian border to prevent illegal crossings."   |
| Comment - 1                        | Most of these people are trying to flee war-torn countries to build a better life...You never know who will become a refugee and when. A little empathy and compassion.   |
| Comment - 2                        | If we blamed those who caused the war instead of blaming these people for fleeing the war and coming to our country, maybe we wouldn't be in this situation today. At this rate, we will start fleeing from Turkey, let's see what Europe will do to us then. |
| Comment - 3                        | When we talk about refugees, don't forget that they also have mothers, fathers, brothers, sisters, children!!! If they are coming over the walls, there is a reason. No one wants to leave their loved ones behind.   |
| Comment - 4                        | Please type refugee camps in Google and look at the images that come up...They know that when they leave their countries they will live in such places at best. Nobody who is not in a difficult situation wants to come into such a life.                    |
| Comment - 5                        | I think you don't understand how desperate it is. Imagine you are on the wing of an airplane moving to escape from your own country. May God not leave anyone in such desperation.  |

## A.2 Questions to Measure Participants' Refugee Attitude

Q1: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?

"Turkey hosts the most significant number of refugees in the world. Therefore, it should not take in any more refugees, and the borders should be closed."

Q2: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?  
"Most refugees who want to come to Turkey are not fleeing war. They come here for economic reasons."

Q3: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?  
"Refugees quickly adapt to Turkey's culture, which increases our cultural diversity."

Q4: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?  
"The judicial crime rate in Turkey has increased due to the growth of the number of refugees."

Q5: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?  
"The opening of new workplaces by refugees helps the Turkish economy to grow."

Q6: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?  
"One of the biggest reasons for the recent housing problems is the refugees coming to our country."

Q7: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?  
"Refugees are one of the reasons why the unemployment rate in Turkey is so high."

Q8: Can you rate the statement on a scale of 1 strongly disagree to 5 strongly agree?  
"Refugees living in Turkey work and pay taxes. Therefore, they should have the same rights as Turkish citizens."

## APPENDIX B

### B.1 Grouped Categories of Respondents' Demographic Characteristics

| Demographic features                      | Codes |
|---|-------|
| Gender                                    |       |
| Women                                     | 1     |
| Men                                       | 2     |
| Age                                       |       |
| 18-29                                     | 1     |
| 30-44                                     | 2     |
| 45-64                                     | 3     |
| 65+                                       | 4     |
| Education level                           |       |
| Graduated from secondary school and below | 1     |
| High school graduate                      | 2     |
| University graduate                       | 3     |
| Master's degree graduate                  | 4     |
| PhD graduate                              | 5     |
| Ethnic identity                           |       |
| Kurds                                     | 1     |
| Turks                                     | 2     |
| Other                                     | 3     |
| Unspecified                               | 4     |
| Voted Party                               |       |
| AKP                                       | 1     |
| CHP                                       | 2     |
| HDP                                       | 3     |
| İyi Party                                 | 4     |
| MHP                                       | 5     |
| Other                                     | 6     |
| Preferred not to say                      | 7     |

## CURRICULUM VITAE

### Personal Information

Name and surname: Gülten Okçuoğlu

### Academic Background

Bachelor's Degree Education

- Kadir Has University, Faculty of Economic, Administrative and Social Sciences  
Department of Economics 2007-2013

Post Graduate Education

- Kadir Has University, School of Graduate Studies, Department of  
Communication Studies, 2020-2022

Foreign Languages: English

### Work Experience

Institutions Served and Their Dates:

- Support Foundation for Civil Society, Communication and Reporting  
Coordinator, 2021 - Current
- APICE- Agenzia di Promozione Integrata per i Cittadini in Europa, Calabria,  
Research Assistance, 2021 - 2021
- TÜBİTAK, Research Assistance, Understanding the dissemination of  
disinformation regarding Covid-19 in Turkey from the media user's perspective  
and developing ideas for preventive measures, 2020-2021